DISSERTATION

Theoretical Study on the Structure and Dynamics of Hydrogen Hydrates

Graduate School of
Natural Science & Technology
Kanazawa University

Division of Mathematical and Physical Sciences

Student ID Number: 1524012015

Name: Yudha Arman

Chief Advisor: Shinichi MIURA, Prof.
Date of Submission: June 28th, 2018

Abstract

Molecular dynamics study of hydrogen hydrates C_1 and C_2 known as filled ice II and Ic, respectively, have been performed. The structure of this two hydrates is closely similar to the respective ice structure encapsulating hydrogen molecule in a lattice site constructed by the framework of hydrogen-bonded water molecules. The hydrates were calculated under temperature condition controlled at 291 K. The pressure of 1 GPa for C_1 and 4 GPa for C_2 was carried out adopting the experimental arrangement where the fast diffusion of hydrogen molecules was observed; here, the H_2:H_2O stoichiometry of 1:6 was used for C_1 and 1:1 for C_2. There are two calculation setups have been carried out, namely full occupation and one vacancy setup, to demonstrate the ideally synthesized and the occupation's defect hydrates, respectively. Longtime simulation has been done to study the diffusion process; here, 100 ns for full occupation and 50 ns for one vacancy setup. In the calculation, many inter-site translations were observed only in the one vacancy setup of both hydrates. This behavior indicates the necessity of occupation's defect to facilitate the diffusion. The calculated hydrogen molecule diffusion coefficient was found to be in good agreement with the experimental results confirming the highly anisotropic diffusion for hydrate C_1 while isotropic for C_2.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays, the long-term mining and usage of unrecoverable fossil energy have made several issues regarding human safety and environment. From the early beginning of production line, the severe impact already noticeable. The spill of the crude product has been reported to give immense implications to surroundings. Ecological problems were generated by the calamity. Also, the release of natural gas in numerous location during exploitation has given more damage to the environment. When concerning the usage, the imperfect combustion systems have released hazardous byproduct. At the early stage of contamination, one would argue that this could be responsible for the several known significant impacts. Eventually, the accumulative disadvantages have expanded massively threatening many life aspects. In several cases, the level content beyond the safety measure of this hazardous byproduct is reported evidence in the air [1, 2]. The critical term is labeled to substances which can implicitly and explicitly be threatening the life-supporting condition. Also, greenhouse effect generated by CO_2 accumulation is now becoming a commonly-heard issue [3]. The fossil fuel burning contributes to the released of CO_2 into the atmosphere.

Less consciousness to make a substitute substance previously shown for an extended period. Since this fossil fuel is the primary source of energy, this unrecoverable substance exploited numerously. As a result, the supply has primarily reduced to be predicted less portion will be left over to sustain the future needs. Concerning this point, awareness to find the substitution candidate has attracted intensive studies. Also, the unwilling effect regarding human safety has given more attention towards the finding. Several aspects are declared to be the necessity during the discovery. The alternative substance should be renewable and efficient. Again, the derivatives produced during the usage should be environmentally clean.

First of all, the review of the replacement candidate is starting at the basic of the carrier itself. There are two known energy carriers, namely electricity and hydrogen. Despite being the lightest and smallest molecule with bond distance is 0.74 Å, the diatomic H_2 have the highest energy content regarding the weight while showing the lowest concerning the volume. Hydrogen also has higher efficiency and easy to store in different ways, thus potentially to be used as a new energy source providing low polluting fuel. This source of energy can be transported and heated in the places where the electricity is hard to

generate. Also, it is practically cheaper to deliver hydrogen with pipeline compared to the long distance wired electricity distribution. When one considering the use of materials encapsulating hydrogen in transporting, efficiency will grow higher. Space shuttle mainly uses hydrogen to produce energy and pure water fuel-cell for the crew.

Hydrogen is bonded into compounds by other molecules. The bonding substance is then can be separated to give H_2 exist in gas phases employing electrolysis, steam reforming, photo-electrolysis and biomass gasification. The respective hydrogen gas can be directly used or be kept in a container to harvest the energy. Compression is needed to maintain hydrogen in the gas phase. The compressed hydrogen, therefore, can be stored in a tube to be delivered or fed into the pipeline network. Given this point, it is advantages to have porous materials which can encapsulate a significant amount of gaseous hydrogen. Again, the ability to be synthesized and kept in ambient pressure and temperature would gain substantial potential benefits. Additionally, to be practical, it is crucial to have host structure able to accommodate hydrogen molecules without having chemical mixtures. Recent studies have proposed several substances to be the proper candidate.

Clathrate hydrate is a crystalline ice material constructed from frozen water molecules accommodating guest molecules. The arrangements assembling cage-like framework in which light noble gaseous or liquid can be encapsulated [4, 5]. The guest molecules which have substantial hydrophobic moiety trapped inside these polyhedral cavities through single and multiple occupations per cage. The encapsulation performed without having the chemical mixture, preventing the host to collapse when positive external pressure is applied [6]. Guest in the gas phase is kept inside cages only by physical bonds. These guest encapsulations have made the resulted structure in a broaden phase-space stability compared to the respective metastable empty structure. These encapsulations were also preventing the cages not to rupture to the typical ice crystal structure or liquid water. This behavior, in the view of the statistical study of several clathrate hydrate systems, were correlated with the van der Waals volume and dipole moment of gaseous species [7]. The strong correlation was shown by the two-phase equilibrium variables, namely dissociation heat and pressure dissociation of ice point. Meanwhile, van der Waals volume of gas species was ranked as the second and third to contribute to the heat and pressure dissociation shifting from unoccupied ice structure, respectively. Compared to the previously discussed variables involved in the reported study, the dipole moment of guest molecule gives lesser influence to the clathrate phase equilibrium. The guest molecules along with hydrogen-water atom released during melting. These remarkable features attract interest towards the study in producing more efficient hydrogen storage materials. To further extent, understanding the physical and dynamic interaction between the foreign molecules and the water framework as host is required.

In the early stage of hydrogen storage research, clathrate hydrates compound were the less known candidate to be proposed and considered to be studied. The flow-line problems occurred in the natural gas exploitation have turned this solid compound into the famously noticeable subject to study; under high pressure and low-temperature, the clathrates built-up formed inside pipeline blocked the gas distribution [8]. Also, Geohazard of the

dissociation of gas hydrate originated from the deep-water deposits have made broaden attention to the clathrate hydrates towards serious interest regarding human safety [9,10]. However, despite the unexpected findings that famously introduce clathrate hydrates, a recent study discovered the potential of this practical host-structure containing gaseous molecules, in particular as the solution of CO_2 removal [11] and hydrogen storage materials [12–16]. The later attracts relentless study to meet the demands of efficient, renewable and clean energy sources. Moreover, theoretical study and experimental findings of hydrogen hydrate revealed the ability of hydrogen to form clathrate hydrate through single and multiple occupancies [17, 18]. This finding against previous believed which stated that inclusion of hydrogen guest was unable to stabilize cage due to its small molecular size. This flawed suggestion of inability due to hydrogen size had spanned for a long time.

There have been several observations noticing advantages from hydrogen hydrate. This material is a non-explosive substance in which essential concerning transportation and handling safety. Also, clean by-products have resulted from the energy harnessing; here, only water and a small concentration of promoters produced. During harnessing, the hydrogen guest is released from the framework through the pressure reduction and small temperature adjustment. This energy carrier stays in its molecular form inside the cage. Moreover, several published data showed that hydrogen hydrate was able to form in the moderate temperature and condition with the need of low promoters concentration; substances such as THF [19, 20], Acetone [21] with low concentration were used to improve operating conditions. These reveal potential features on having hydrogen storage economically with less energy effort in controlling synthesize process. There also numerous hydrogen hydrate having higher hydrogen content per unit mass and volume had been discovered. [12, 13, 22, 23]

However, several challenges had been experienced to make outstanding efforts on having clathrate hydrate as hydrogen storage. The duration of time to synthesize the compound featured in several hydrogen hydrates due to the reluctant of formation kinetic obligate effort to show less effectiveness. Stability issue at ambient thermodynamic condition showed similar high effort lineaments of hydrogen hydrate. Also, when one include promoters to have a better operating condition, storage capacity will be reduced since these molecules occupying the available spaces. Many studies have been conducted to overcome the limitation. However, a deduction of one aspect should be made when gaining the other [24].

Hydrogen gas can also be found encapsulated in the cage-like water framework without the need of promoters to support. For instance, high amount of hydrogen gas can be stored in lattice sites of ice II and Ic, with H_2: H_2O stoichiometry is 1:6 and 1:1, respectively [23]. Instead of clathrate hydrate, these two known as filled ices of C_1 and C_2, respectively. The inclusion stabilizes the ice structure in an extended range of pressure and temperature compared to the unfilled composition [25–29]. Hydrate C_1 is firmly having ice II structure while hydrate C_2 shows cubic ice framework. Not only hydrogen, other light noble gas molecules can also be occupied [30, 31]. However, the high-pressure condition is needed when synthesizing the materials to compensate the interaction energy between guest and

surrounding host molecules which shows repulsive character due to the small size of lattice site. The guests themselves are needed to stabilize the host-framework structure. The high ratio between host-guest is a necessity to avoid lattice site's rupture when highly needed pressure is applied to the system.

The crystal structure of hydrogen hydrate C_1 is known to be rhombohedral in the $R\bar{3}$ space group. It can also be viewed as the hexagonal unit cell by using Miller indices relation $-h+k+l=3n$ where n is an integer. Under 2.1 GPa and 295(3) K, this hydrate was reported to have a hexagonal unit cell containing 36 water molecules [23]. The full arrangement consists of the hexagonal array of hydrogen-bonded water molecules lies on the surface known as the basal plane. The planes are piled along the optical c axis to form hexagonal tube-like columns in a similar way as the ice II.

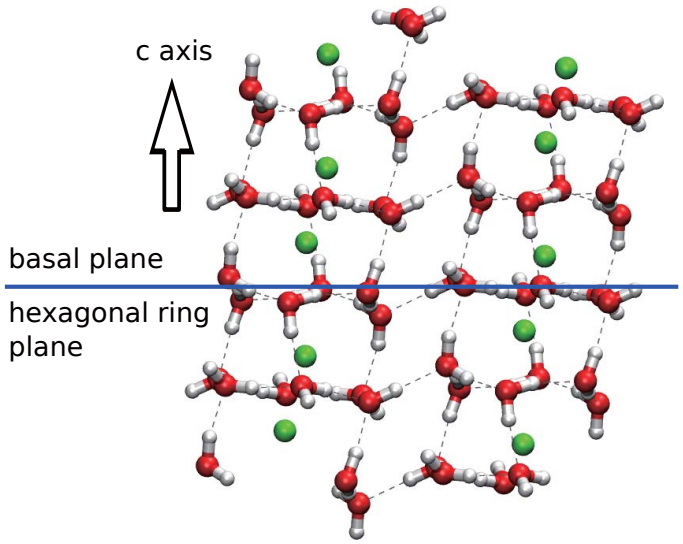


Fig. 1.1. Molecule model for C_1 system. Green spheres describe hydrogen guest.

On the other hand, under 3.1 GPa of pressure, hydrogen hydrate C_2 was reported to have a cubic crystal structure [23]. The arrangement of water molecules gives eight faces forming diamond-like crystal structures. Surfaces containing hexagonal shaped of hydrogen-bonded water molecules are formed by stacking (111) faces in a disordered way along (110) direction. This (111) plane is similar to the unoccupied cubic ice structure and perfectly coherent to the basal plane of hexagonal ice [32]. The hydrate was able to store maximum 10%wt H_2 capacity equivalent to 1:1 molar ratio between hydrogen and host water framework. The stoichiometry gives C_2 characterized as a hydrogen-rich material compound, thus suitable for hydrogen storage.

Fast diffusion of hydrogen guest molecules has been reported in these dense hydrates under highly compressed condition [33,34]. The observed rapid hydrogen molecules inter-site translation is comparable to the molecular diffusion in a liquid state with less affected by the pressure. Moreover, concerning the diffusion trajectory, anisotropic is suggested in hydrate C_1 while isotropic is for hydrate C_2. This suggestion was governed from more to the structural arrangement's consideration rather than theoretically observed. The details

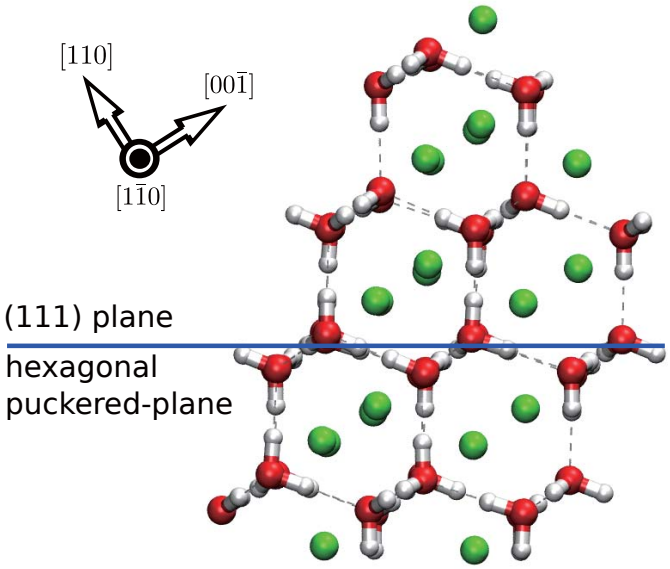


Fig. 1.2. Molecule model for C_2 system. Green spheres describe hydrogen guest.

mechanism could be crucial in designing better hydrogen storage in these filled ices. There could be a factor to drive these guests diffusion among polyhedral cavity.

There are several fascinating aspects regarding molecular encapsulation in ice-resembling structure. In clathrate hydrate systems, the guest to diffuse is placed in polyhedral cavities of ice structure through weak Van der Waals interaction. Correspondingly, no chemical mixture is observed. The dimension size of the polyhedral enclosure is reported to give influence in the number of occupation and the guest mobility. It is therefore needed to have a microscopic understanding of this interaction for several applications. In the development of hydrogen storage, numerous study gives evidence on the importance of structural size in the intra-cage dynamics [35]. Moreover, in the specific system such as filled ices, the small size of lattice site give high repulsion characteristic shown by the guest-host interaction [29]. Therefore, high pressure is needed to compensate the confinement. As a consequence of the pressurization, more dynamic is expected to be observed since the density of guest molecule is denser compared to its condensed state. Imperfect confinement structure also influenced the mobility of the confined species. Inter-site translation is observed in clathrate hydrate by this kind of water vacancy [36, 37]. With the defect, free energy of lattice site's boundary is lowering to allow the guest to diffuse.

1.1 Purpose

This research is performed to theoretically study the dynamical properties occurred in hydrogen-filled ice II and Ic. Both systems can be considered as a hydrogen-rich material compound where rapid diffusion of the guest was found with less influenced by the pressure. It has been suggested that hydrogen molecules are denser compared to their

condensed phase under this condition. This dynamics-related structure-governing kinetics is useful to be studied. They determine the duration time of synthesizing and the dissociation process. The same rapid diffusion of guest molecules should be captured in the calculation to justify the provided evidence of microscopic details of hydrogen molecule's dynamic in this study.

In this doctoral thesis document, the report of the molecular dynamics (MD) calculation result of hydrogen hydrate C_1 and C_2 will be presented. There are several confirmation points of the previous observations in C_1 and C_2 structures and relevant additional information of diffusion mechanism provided to be done. Also, details of a long time MD calculation when investigating the inter-site diffusion of the encapsulated hydrogen molecules is discussed. There were two different setups on each of the filled ices structure. Systems with the maximum load of hydrogen guest molecules were the first system to simulate while single lattice site left unoccupied be the second. These setups were delivered based on the results of a vacancy-driven diffusion that had been studied extensively in numerous clathrate hydrate systems [37–39]. The result was suggesting the necessity of an imperfect guest occupation in ice structure to facilitate the diffusion process.

1.2 Dissertation Organization

This dissertation is organized as follows. In Chapter 2, the Ice and Clathrate Hydrtaes are discussed. The method when performing molecular dynamics calculation are briefly presented in Chapter 3. In Chapter 4, calculation results and post-production analysis are discussed. Concluding remarks discussing summary and future works are presented in Chapter 5. The appendix is used to explain the mathematical background of the method utilized during the calculation. Also, in this appendix, post-production analysis is discussed and divided into several sections. To be more specific, the brief discussion of molecular dynamics method is presented in Chapter A while in Chapter B for the diffusion coefficient code listing. Potential energy map and IR spectra calculation codes are presented in Chapter C and Chapter D, respectively. All structure figures presented in this report are produced by using Vesta [40] and VMD [41]. Again, the calculations of hydrogen bond energy and distance were performed with the help of VMD program.

Chapter 2

Ice and Clathrate Hydrate

2.1 Ice

Ice is a solid phase of frozen water, formed by tetrahedrally coordinated of water molecules connected by the hydrogen bond to the four others. The angle of this tetrahedral connection equal to 109.47° which is slightly higher compares to the H-O-H angle. Concerning oxygen positions relative to other oxygen water molecules, there are eight known crystalline phases forming the hexagonal lattice of ice to have ordered and disordered arrangements of hydrogen water depending on water and pressure. Meanwhile, two amorphous phases (non-crystals) of ice are identified since no long-range ordered arrangements featured. This arrangement can be the cause of insufficient time to form a crystal lattice when rapid cooling is applied. Meanwhile, decompression of low-temperature ice can give the similar amorphous ice structure.

On earth surface and atmosphere, all of ice structured as hexagonal crystalline and also known by its arrangements abbreviation, namely ice Ih. The phase transition of water to ice Ih is commonly achieved by cooling liquid water to about 273.15 K at ambient pressure. Similar to what happened in the forming of frost formation, the transition can also occur directly by depositing water vapor. Symmetry is preserved by water molecules obeying ice rules [42] where the arrangements are allowing total dipole moment close to zero. When the rules are disobeyed, the ice crystal would be weakened by the failure of neighboring water adjusting the hydrogen bond to prohibit molecular rotation. The same charged atom would be in line without having intervening of the opposite in between. This weakened is known as Bjerrum defect [43]. Also, the ionic defect may occur when H_3O^+ and OH^- ions are present.

The many variants of ice are possible to observe since ice Ih featured open structure of tetrahedrally arranged water molecules in which crushed dynamically when high pressure is applied. The pressure shorten the distances between the second shell neighbors thus reducing ice volume while gaining van der Waals effects. Since there is an anomalous change in entropy with volume and the negative expansion coefficient of water, Clausius Clapeyron equation is adapted in many ice phase changes.

2.1.1 Ice II

Ice II has relative permittivity or dielectric constant of about 3.7. This crystalline water arrangement has triple points with other forms of ices. Firstly, under temperature and pressure condition of -34.7 °C and 212.9 MPa, respectively, Ice II coincide with hexagonal ice and ice-three. Secondly, ice II coexist with ice-three and ice-five at -24.3 ° of temperature and 344.3 MPa of pressure. Thirdly, ice-eleven and ice Ih have existed side by side to this ice II under -199.8 °C of temperature and 70 MPa of pressure. And finally, ice-six and ice V can also be found along with ice II estimated at -55 °C and 620 MPa.

Correspond to this phase-space diagram, increasing the temperature towards the melting point of Ice II will turn this form of ice into hexagonal ice, Ice III, V and VI. At ambient pressure and temperature above 160 K, this ice turned into ice Ic. However, when 0.5 mol of NH_4^+ F^- is added, this ice would be disappeared from the water phase diagram [44]. Furthermore, this form of ice is at its metastable state when conditioned below 100 K in the pressure range of ambient to 5 GPa.

Bringing hexagonal ice into a thermodynamic condition of 198 K temperature and pressure of 300 MPa will collapse half of its hexagonal channel to form ice II. This new arrangement has the detached ice Ih hexagonal channel connected by more compact hydrogen bond. The density is then about 1.16 g cm^{-3}. The deformation from the Ice Ih is detailed as follow. The hexameric ring featured in ice Ih is rotated to the right angle of 30 ° horizontally on the basal plane. Then, the rotated ring is slightly moved vertically and having hydrogen bonded to each other. The connections between ice Ih and ice II can also be explained as follow. The single crystal of Ice Ih can perform single crystal of ice II or twin ice II. This transformation can be done by rotating 180 ° the c axis channel concerning each other.

Another way to synthesize Ice II is by decompressing ice V at 238 K. This is achieved at ambient pressure. It is also possible to make ice II from ice III by temperature reduction, although it is difficult to perform.

Hydrogen bonds observed from ice II are ordered and fixed. Each water molecules bonded to other four by means two as the donor while the other two as acceptor. In real life, there is no ice II arrangement to have disordered hydrogen bond. However, the theoretical study revealed possibilities to have partially disordered proton configuration found in ice II [45]. Since some of them can be seen bending due to less varying of isolated water molecules, the strength is thus weaker compared to the hexagonal ice. Despite preserving half of the hexagonal rings featured by ice Ih, the transition inherits 3/4 of total hydrogen bond performing puckered and flat rings layers change repeatedly. The remaining 1/3 at this hexagonal channel constitute from Ice Ih are conserved. The newly 1/6 at this channel thus give ordered arrangements.

The unit cell of ice II consist of 12 water molecules, with lattice parameter $a = 7.78$ Å and have α, β and γ equal to 113 ° to form Rhombohedral R_3 crystal [46]. There is another way to visualize ice II by considering other unit cells of 36 water molecules with lattice parameters a, b, and $c = 12.935$, 12.935 , 6.233 , respectively. This structure was

found under 225 K of temperature and 0.25 GPa of pressure [47]. It is a lot simple to visualize and a lot easier to interpret since the crystal is hexagonal.

2.1.2 Ice Ic

Cubic ice, also known as Ice Ic is found in the metastable state. Within time, this form of ice will be transformed into hexagonal ice through the increasing disordered of the stacking-disordered ice. Sometimes transitional state between hexagonal and cubic ice can be found depending on the formation and history of the cubic ice. Compared to ice Ih, ice Ic has lower melting pressure while showing higher vapor pressure. Cubic ice can be found in the upper atmosphere and be freezing in confinement or porous aqueous systems.

Ice Ic formed by condensation of water vapor at ambient pressure and low temperature less than -80 °C. It can also be created when water droplet with the diameter about 6 μm is cooled in -38 °C. Lower interfacial free energy is found in water droplets which have a radius smaller than 15 nm at 160-220 K. This shape is the most preferred phase space to form cubic ice. By reducing pressure on the high-pressure ice at 77 K, cubic ice crystal can also be synthesized [48]. Homogeneous ice in profoundly supercooled nano drops (r $\approx$ 10 nm) of water at about 225 K would crystallize to have 78 % cubic ice and 22 % hexagonal ice, thus effectively 44 % of stacking-disordered ice can be found. Stacking-disordered always found when ice crystallizes from supercooled water. It is difficult to have a cubic structure with a higher percentage in a cubic ice structure even though cubic ice always used to acknowledge this not pure ice crystal. Heat evolution up to 50 K mol^{-1} is needed to transform ice Ih into ice Ic irreversibly in 170-220 K, but the process took very long time.

Similar to ice Ih in having moderate open low-density structure, fcc crystal structure of ice Ic showed less packing density compared to simple cubic, or face-centered cubic would typically have. However, it's density almost similar to ice Ih since more staggered settlements of the hydrogen bond concerning all neighbors are shown by ice Ic, rather than to the fcc packing density. Another structure difference demonstrated by ice Ic compared to ice Ih is in the second neighbor arrangements.

The cubic crystal has unit cell dimension is 6.358 with α, β, and γ angles are equal to 90°. Each unit cell contain 8 water molecules [49]. Water molecules are arranged to obey ice rules [42] having two hydrogen atoms are placed near oxygen while one hydrogen atom is intervening oxygen-oxygen bond. Also, above 5 K, protons can move between water molecules [50]. All of the hydrogen bonds are disordered and not straight since the angle of isolated water molecule arranged to have four nearest tetrahedrally structured neighbors does not much modified. There are 12 tetrahedral arrangements of water molecules at the second neighbors similar to ice Ih. Less water molecule is found in ice Ic in the second hydration shells compared to ice Ih to cause less stable arrangements. This arrangement also gives cage-like structure has smaller volume compared to the ice Ih.

As can be found in ice Ih and low-density amorphous ice while showing different behavior compared to most crystals, thermal conductivity anomaly was found to reduce when

pressure is increased in ice Ic. This behavior is due to the decreasing of transverse sound velocity caused by the changing of hydrogen bonding [51].

2.2 Clathrate Hydrate

Clathrate hydrate is the name to acknowledge compounds composed by firmly packed-frozen water molecules framework accommodating gaseous molecules. This resembling-ice structure is governed by the arrangements of hydrogen-bonded water molecules in such a way its oxygen atom becomes a point-edge of four cage-like structure. Some of the polyhedral cavity may or may not be occupied. The non-stoichiometric molecule accommodated by the cage also known as the guest. It is usually stored in its molecular form at gas state. The molecular inclusion involving neither chemical mixture nor covalent bonding. It is observed that there are van der Waals interactions between guest molecules and it's surrounding cage. At the early stage of the discovery, hydrogen bonding is aligned disordered imperfectly and obeying ice rules [42]. This compound is in the same equilibrium condition as its engaged guest. Meanwhile, the cavity occupation number depends on temperature and pressure. The solubility is higher in their solid phase rather than their liquid state. Interestingly, this compound is in the unstable state at positive pressure without the proper amount of guest filling their empty void.

The hollow cage-like structure would easily rupture due to applied pressure. Subsequently, when continues breakage begin to show the melts of clathrate ice, this cage-like structure will slowly dissipate afterward [52]. Generated universal model of clathrate hydrate occupying guest was simulated to reveal that this compound starts dissociating into surface ice and guest gas below their melting point [53]. Meanwhile, NMR study showed that there is a significant influence made by the solid phase clathrate hydrate to its surrounding melted solution during formation and dissociation [54].

As previously described, problems reported on offshore oil exploitation when clogging in the flow-line caused by clathrate hydrate built-up has given this solid compounds their first academic introduction [8]. Also, massive release of their hazardous inclusion species from deep-water deposits [9, 10] provide further attention to outset relentless study to reveal potential usage as energy storage [12–16] and the global warming gas removal [11].

Clathrate hydrate can be found accommodating one species of gas. However, it is often observed that the inclusion constitute more than one gas component to stabilize cages structure at ambient pressure and temperature, with the exception found on bromide clathrate [55]. Above 270 K and ambient pressure, clathrate hydrates will release their guest molecules. However, at the same pressure with lower temperature, they are at metastable state for several hours. Complex phase behaviors were shown by pressurization under high-pressure condition. In this thermodynamic state, smaller guest species was possible to be found isolated in hexagonal structured ice. Structure breakdown to form denser ice accompanied by solid-phase guest molecules were observed when pressurization continues.

There are three known structure types for clathrate hydrates, namely sI, sII and hexagonal known as sH. The two first mentioned have the cubic crystalline structure of Pm_3n and Fd_3m, respectively, while the last has hexagonal based structure. CO_2 and CH_4 hydrates have sI structure, O_2 and N_2 have sII and some solid-phase materials found in slurry reservoir fluids are suggested to have sH structure coexist with sII [56].

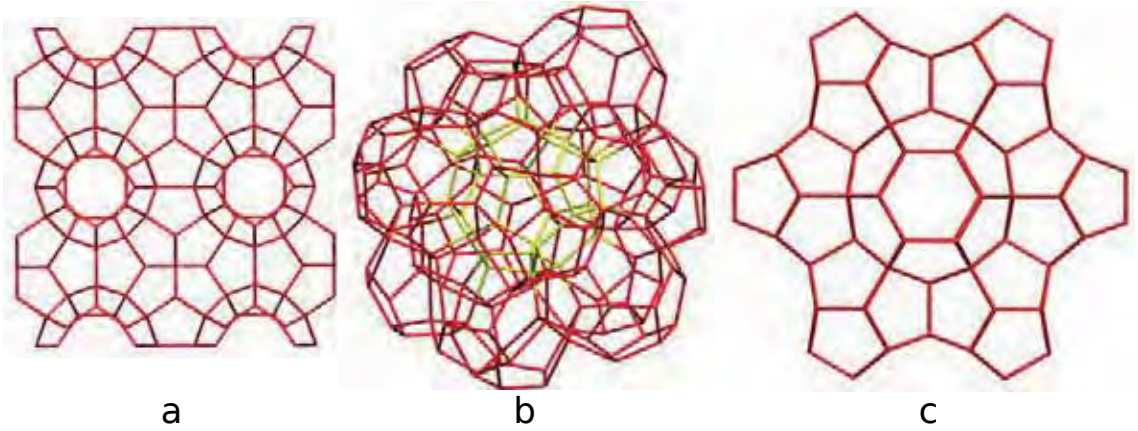


Fig. 2.1. Clathrate hydrate structure of (a) sI, (b) sII and (c) sH.

The unit cell of sI containing 46 water molecules in a unit cell. Each unit cell then tiled to form 3 dimensional structure of small-pentagonal-dodecahedron cages (5^{12}) and large-tetradecahedron cages which are hexagonal truncated trapezohedron ($5^{12}6^2$). The two type of enclosures are arranged to show WeairePhelan structure with open tetrahedral O-O-O connection to deviate about 3.7°. The average distance between oxygen-water atoms is 2.793 Å. Meanwhile, a lot more water molecules in a unit cell are needed to form the sII structure; here, a unit cell is containing 136 water molecules with the O-O average distance is 2.790 Å.

Unit cells are tiled to form sixteen similar small-pentagonal-dodecahedron (5^{12}) correspondingly as sI. Meanwhile, the arrangement also gives eight large hexahedra ($5^{12}6^4$) cages in which 3° angle deviation of tetrahedrally connected oxygen water to its adjacent neighbor is shown. Three types of enclosures with the composition of three small 5^12 cages, two small $4^35^66^3$ cages and one large $5^{12}6^8$ cage are arranged in which 34 water molecules are found in a unit cell to give sH structure. As previously mentioned, the large cage in this hexagonal structure can be filled by large molecules such as butane and hydrocarbon. The most stable structures are shown by sI and II.

Continues discovery of clathrate hydrates revealed more structure variants in which some would relate to Frank-Kasper (FK) structure [57]. The first variant is clathrate with tetragonal s-T structure. This structure has 12 water molecules in a unit cell arranged to make cells of $4^25^86^4$ cavities [58]. The second is clathrate materials having cavities such as 6×5^{12}, $4\times5^{12}6^2$ and $4\times5^{12}6^3$. There are 80 water molecules are found in a unit cell. This structure also known as the tetragonal s-K structure and can be found in Bromine hydrate [59]. The third variant to relate with FK structure is cubic s-III clathrate hydrate [60]. The structure have $2\times4^{12}6^88^6$ and $6\times4^88^2$ cavities built from

water molecules arrangements; here, a unit cell have 48 water molecules.

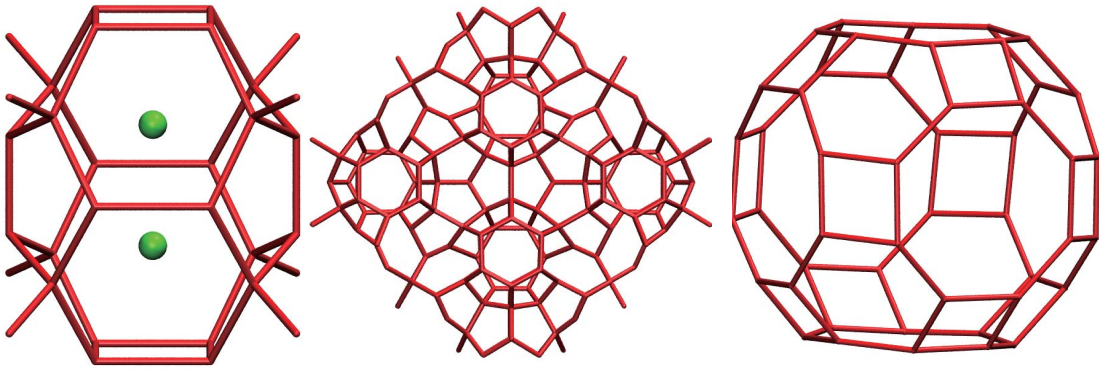


Fig. 2.2. Clathrate hydrate structure of tetragonal (a) s - T [58], (b) s - K [59] and (c) s - III [60]. Green spheres indicates guest molecules.

2.3 Filled Ices

Molecular hydrogen can form clathrate hydrate through single and multiple occupancies to fit in each cavity. Under high pressure, this small guest molecules can be occupied in different host arrangements rather than known clathrate hydrate structures; here, the structure mostly known as filled ice. The composition of this hydrogen hydrate material under 400 MPa pressure and 280 K temperature is $(H_2O)_2H_2$ where ice structure framework observed as host. Filled ice from different guest inclusion can also be constructed by compressing clathrate hydrate. For instance, CO_2-filled ice is originated from sI CO_2 clathrate hydrate [61]. Meanwhile, filled ice where ice-XVII is host framework can be observed by compressing sII CO_2 clathrate structure [62]. Detailed schematic diagram regarding this phase transition concerning pressure at room temperature was presented in [63].

There are several filled ice structures extensively studied, namely C_1 hydrate which has ice II host structure to encapsulate H_2 and C_2 hydrate system occupying the same small-sized hydrogen gas in lattice site of ice Ic water framework [23]. These two have H_2: H_2O stoichiometry of 1:6 for C_1 while 1:1 shown by C_2 to be considered as a hydrogen-rich compound. In the case of C_2 system, pressure dependence revealed larger than the unoccupied would show under the same pressure condition; similar behavior affected on ν_{OH} frequency and d_{OH} distance of unfilled ice Ic structure would be under the twice higher pressure condition [25]. Meanwhile, since H_2+ H_2O is a binary system, C_1 and C_2 were found to coexist under pressure condition ranged from 2.3 GPa to 3.1 GPa.

Furthermore, MH-III structure hosting methane was found at about 2 GPa in which water molecules are arranged very similar to hexagonal ice structure [64]; this methane-water compound is considered as the first methane guest encapsulated in hydrate structure known as filled ice. Similar to hexagonal ice, hydrogen bond arrangements revealed in

this hydrate are reasonably disordered in Imcm space group, where a = 4.7458(5) Å, b = 8.0644(9) Å and c = 7.8453(7) Å. The methane-guest shows strong repulsion character with regards to the C-O distances and showed increases strongly with pressure when the system at 11 GPa. This repulsive character already observed at 3.0 GPa with interaction energy in the order of 1 kJ/mol [65].

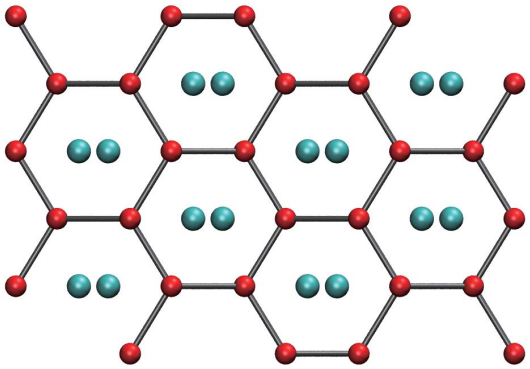


Fig. 2.3. MH-III structure [64].

Another structure, namely C_0 was recently discovered. However, the definite structure of this C_0 hydrate system had not been established. Previous observation at 80 K and 0.5 GPa pressure employing ex-situ X-ray diffraction proposed structure of space group $P3_112$ (or $P3_121$) in which water molecules presence in sites 3a2 with half occupancy [66]. However, this spiral H-bond network of trigonal arrangements showed a violation of ice rules. Subsequently, Strobel et al. [67] continues the effort to observe properly the C_0 structure utilizing XRD suggesting two structures, tetragonal S-T structure and α - and β - quartz arrangements. Next, Smirnov et al. [68] revealed that the partially occupied sites were N_2 molecules included during recovery while H_2 occupying the channel; this configuration is later known as C_0-II. The result also indicates that this C_0 have tetragonal S-T structure. Recent XRD study revealed that the C_0 have structure closely similar to Smirnov et al. [68] with hexagonal symmetry [69]. Recent findings stated that at 100-270 K and pressures range from 360 MPa to 700 MPa [70], C_0 have stability region at the sII clathrate and C_1 based ice II structure. The results revealed hexagonal symmetry with spacegroup of $P6_122$ when this unoccupied C_0 system recovered to ambient pressure; this new form of hexagonal ice structure named as ice XVII. Proton-ordering and guest occupancies later determined by in-situ neutron and X-ray diffraction. Along with CO_2 hydrate-II, H_2 hydrate of C_0 show similar structure of chiral and has large open spiral channels for the guest to move [61] freely. This $P6_122$ spacegroup have a = 6.2763(5) Å and c = 6.1875(19) Å with density of 0.991 g/cm^3, exhibits new features of hydrate and hydrogen bonded was suggested disordered.

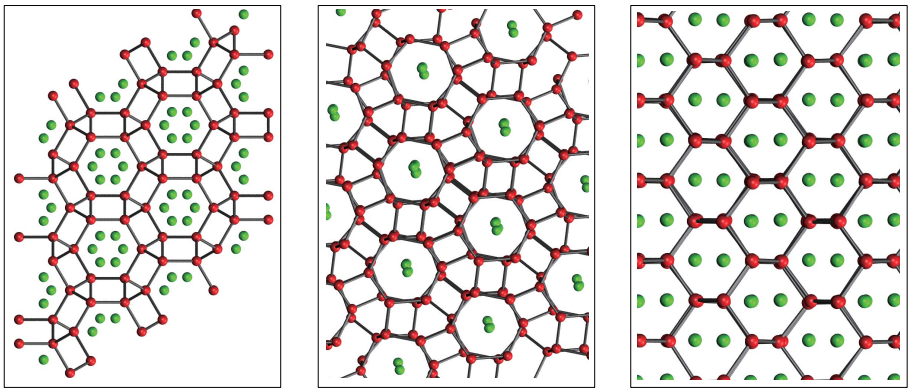


Fig. 2.4. Filled ice structure of (a) C_0 [61] (b) C_1 and (c) C_2. Red and Green spheres indicate oxygen atoms and guest molecules, respectively. Grey line describes hydrogen bond connecting water frameworks. C_1 and C_2 systems are redraw with ice configurations taken from Buch et al. [71] while guest molecules are included without energy minimization. Both structures are viewed along hexagonal puckered-plane

Chapter 3

Molecular Dynamics Calculation

Molecular dynamics (MD) calculation were performed to theoretically study hydrogen hydrate of filled ice II (C_1) and Ic (C_2). Long-time calculation of 100 ns was needed, in particular, to investigate the hydrogen molecules diffusion, in the full occupation setup. Meanwhile, 50 ns were performed for the one vacancy setup. MD calculations were done by using LAMMPS package [72]. The usage starts from energy minimization to the production run. The simulated system was conditioned to reproduce the previous experimental setups. In this condition, fast diffusion of hydrogen guest molecules is observed. The rapid diffusion of the guest showed less influenced by the applied high pressure within a certain range.

When simulating hydrate C_1 system, 1 GPa of pressure and 291 K of temperature were controlled during equilibration and production run. The frozen-water molecules framework has $R\bar{3}$ space group structure taken from Busch et al. [71] satisfying Bernal-Fowler rules [42]. The structure has rhombohedral unit cell with lattice parameter a= 7.78 Å while α, β, and γ is 113.1 ° containing 12 water molecules. Meanwhile, the calculation was performed with 3 × 3 × 3 unit cells containing 324 water molecules and encapsulating 54 hydrogen guest molecules. Initially, the guest was placed in the lattice site. The H_2: H_2O stoichiometry was following previous experimental results of Vos et al. [23] for full occupation. Imperfect synthesized result by giving one lattice site unoccupied was prepared to study occupation defect-driven diffusion; here, this setup was named as one vacancy system. These two distinctive confinement setups were then

Table 3.1. The potential parameter for water and hydrogen guest molecules used in MD calculation. For H_2 guest molecule, both LJ and charge site are located at the same center of mass. Meanwhile, Oxygen atom's position of water molecule is used as LJ calculation point while the negative charge placed at M position, located between the center of H-H and Oxygen.

Model	Site	q(|e|)	σ(Å)	ε(kcal/mol)
H_2 [73]	Center	0.0	2.96	0.067962
TIP4P/Ice [74]	H	0.5897		
	O		3.1668	0.21084
	M	-1.1794		

Table 3.2. Host Framework Setup

Crystal	C_1 (Rhombohedral)	C_2 (Cubic)
Unit cells	$3 \times 3 \times 3$	$3 \times 3 \times 3$
H_2O / unit cell	12	8
a,b,c (Å)	7.78	6.358
α, β, γ	113.1°	90°

entering energy minimization employing conjugate gradient methods. During this process, the water molecules and guest's position were corrected concerning their pair-interaction to achieve minimum energy arrangements. The observed cavity volume was fluctuated slightly associating to the guest occupation. The resulted arrangements were then entering equilibration stage under the thermodynamic conditions previously described. During equilibration run of 50 ps, energy conservation and thermodynamic fluctuations were carefully observed. Finally, after the completion of the previous stage, production run was performed in which 100 ns for full occupation setup. Although longer calculation was performed, no inter-site diffusion was achieved in full occupation system eventually. Meanwhile, 50 ns was utilized on one vacancy system; here, many inter-site translations performed by hydrogen molecules are already observed.

In a closely similar treatment, calculation of hydrate C_2 system adopted the same reference of the experimental result as hydrate C_1, in which 4 GPa of pressure and 291 K of temperature. With regards to the arrangement, ice Ic structure was also taken from Buch et al. [71]. The unit cell is cubic with lattice parameter a=7.78 Å composed of eight water molecules. When performing the calculation, ice Ic structure of $3 \times 3 \times 3$ unit cells containing 216 water molecules were used as the host to encapsulate 54 H_2 molecules. One vacancy setup also employed in this hydrate with the similar purpose as filled ice II. The guest positions were corrected during minimization process along with lattice site of water molecules framework. Minimization, equilibration and production calculations also treated in the same manner as hydrate C_1.

During calculation of all setups and hydrates, thermodynamics conditions were controlled by using LAMMPS fix npt styles. This fix styles uses Verlet and rRESPA time integrators [75] on Nosè-Hoover [76, 77] form of Shinoda et al. [78] equation of motion. The schemes combined the hydrostatic equations [79] with the strain energy [80]. The desired temperature and pressure were highly preserved by implementing Nosé- Hoover chain method. A chain of 4 thermostats was coupled to the system particles, while the other four to the barostat variables. All of the previously mentioned algorithms will be described further in the appendix.

Meanwhile, to simulate the intermolecular interaction, the sum of Lennard-Jones (LJ) pair potentials and the Coulomb defined by

$$\phi = \sum_{i,j} \left[4\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_i q_j e^2}{r_{ij}} \right], \tag{3.1}$$

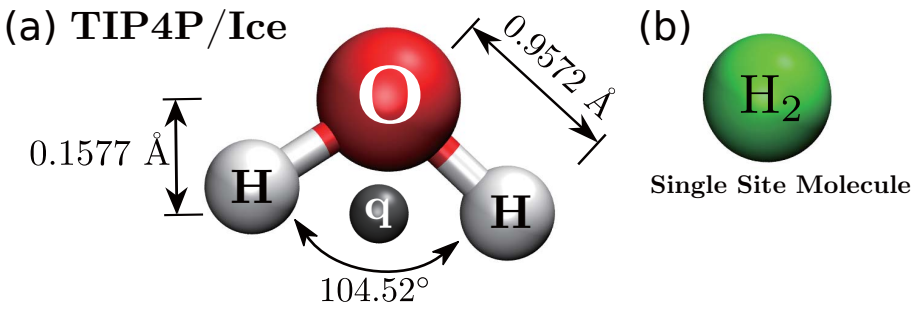


Fig. 3.1. Molecule model for (a) water framework, and (b) guest molecule

was used. To calculate the electrostatic interaction energy contribution, long-range Coulomb interaction was evaluated by using a particle-particle particle-mesh Ewald method [81] method with tail correction on LJ pair potential. The cut-off radius of 10 Å was used for the pair energy interaction. To describe the guest molecules, one site H_2 model was used. The model was proposed by [73] which shows the superfluidity of LJ (12,6) in finite-sized ultracold H_2 bosonic clusters. To model host-water framework, TIP4P/Ice [74] was used. This water model is suitable to be used in solid-phase compound calculation with better reckoning ice properties compared to the TIP4P. To make the water model rigid, SHAKE algorithm [82] was used in preserving covalent bond between Oxygen and Hydrogen atom of water molecules to a distance at 0.9572 Å; here, angle was also preserve to 104.52°. Further discussion of the previously mentioned method are presented in the appendix.

When calculating pair-potential energy with equation 3.1, parameters that were used are listed in 3.1. Meanwhile, the parameter of intermolecular interaction between water and hydrogen guest molecules was calculated by using Lorentz-Berthelot [83, 84] combination rules, written by:

$$\sigma_{12} = \frac{\sigma_1 + \sigma_2}{2} \tag{3.2}$$

$$\varepsilon_{12} = \sqrt{\varepsilon_1 \varepsilon_2} \ . \tag{3.3}$$

A single site of hydrogen guest molecules was used during the calculation of all systems. The potential parameters were taken from Sevryuk et al. [73] Details of this system's setup are listed in table 3.2. For the details of thermodynamic condition, table 3.3 collected the value used in the calculation along with guest setup information on every simulated system. When performing minimization, 1×10^{-25} of relative energy differences was used as the stopping criteria between the consecutive iteration. Meanwhile, in the equilibrium and production run, thermostat and barostat were implemented with 0.1 ps and 1.0 ps of relaxation time respectively. Time step of 1 fs is used for the production run for all simulated systems while for the van der Waals cut off radius is 10.0 Å. Periodic boundary condition was used on each simulated systems to overcome the surface effect. In appendix A, a scheme to handle periodic boundary condition is explained since non-orthogonal box

Table 3.3. Hydrogen molecules and NPT Condition

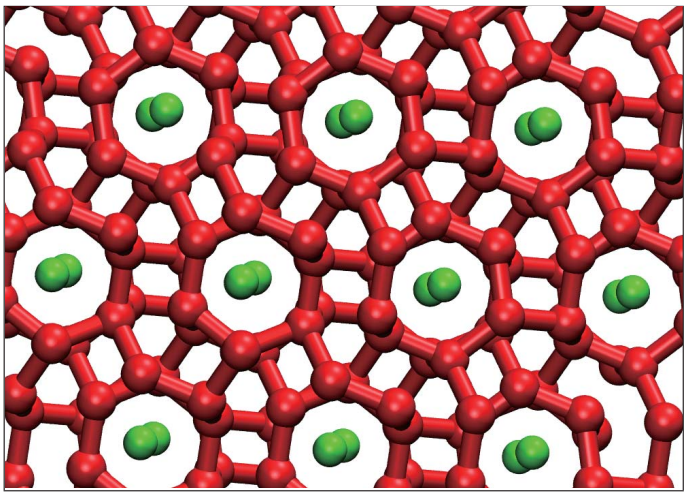
Conf.		H_2O	H_2	P (GPa)	T (K)
C_1	Full	324	54	1	291
	One Vacancy System		53		
C_2	Full	216	216	4	
	One Vacancy System		215		

setup was used when simulating hydrogen hydrate C_1 system. The scheme can be used for both orthogonal and non-orthogonal axis system.

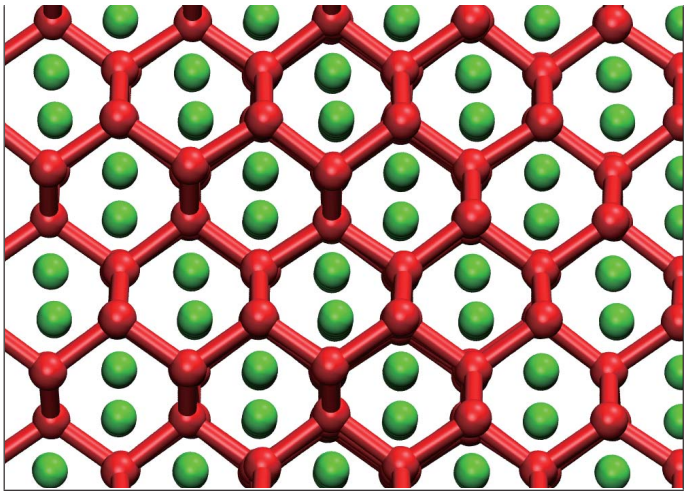
To calculate the diffusion coefficient, the following equation

$$\langle |\Delta \mathbf{r}(t)|^2 \rangle = 2nDt, \tag{3.4}$$

was used where $\langle |\Delta \mathbf{r}(t)|^2 \rangle$ is the mean square displacement of all hydrogen guest molecules, D is diffusion coefficient, n is dimensionality and t is correlation time. In this present study, n equal to 3 was used for all diffusion coefficient calculation. When calculating diffusion coefficient in one vacancy systems, 50 ns of calculation trajectory data was used for 2 ns of correlation time. When calculating this property, a trajectory of hydrogen molecules contributed to the guest diffusion coefficient were secluded from other molecule exist in the system. Periodic boundary condition was employed closely followed the scheme listed in Appendix A. Linear regression was performed to calculate tangent of the plot of mean square displacement. As can be seen from Fig. 3.2 (a), the arrangement of water molecules in hydrate C_1 and C_2 consist of lattice sites assembled in such a way to have different neighboring distance concerning the crystallographic structure. In the case of hydrate C_1, the nearest separation along optical c axis is three times shorter compared to the separation along the basal plane. The remarkable difference in adjacent hydrogen molecules between the two crystallographic axes suggested the highly anisotropic diffusion would occur in this hydrate. Therefore, only hydrogen molecules located along a column of optical axis direction were used in diffusion coefficient calculation. In the case of hydrate C_2, hydrogen molecules are arranged to be tetrahedrally connected. This arrangement is similar to the tetrahedrally connected water molecules in which forming hexagonal puckered-plane. In a similar manner to the arrangement of oxygen atom in the water molecule's framework, hydrogen molecules positioned at a vertex of a hexagonal ring where the distance between adjacent vertices is equal. Within this structure, all of the hydrogen molecules are therefore used in the calculation of diffusion coefficient. The neighbor list was updated every 1 calculation step with extra distance beyond force cutoff was 2 Å. In this calculation, a neighbor-list build can be performed if some atom traveled a distance of more than half of its neighbor's skin value at the last build.



(a)



(b)

Fig. 3.2. The arrangement of oxygen atoms (red spheres) and H_2 molecules (green spheres) in (a) hydrate C_1 and (b) C_2. Hydrate C_1 is viewed parallel to the basal plane and C_2 on (110) plane. Bond connecting the adjacent oxygen atoms and hydrogen molecules are drawn to clarify the structural configuration with the distance criteria 3.0 Å for hydrate C_1 and 2.8 Å for C_2.

Chapter 4

Results and Discussion

4.1 Equilibrium Condition

First of all, evaluation of equilibrium condition from all of the simulated systems is presented. When doing this evaluation, the following analysis will further describe controlled thermodynamic, and energy resulted from MD calculation concerning calculation-time. The analysis was taken in the production-run stage of the two different calculation time. For time-series plot of total and potential energy, the result will indicate whether the calculation is in the equilibrium state or necessarily need more equilibration time. Meanwhile, for the controlled thermodynamic properties such as temperature and pressure, the plot will mainly describe whether the average value coincide with the desired condition properly or the contrary. The evaluation will be presented on every simulated system, namely C_1 of full occupation and one vacancy systems, and C_2 system with two similar setups of lattice site's occupation as in the C_1. Plot of controlled thermodynamics and energy are given in Figs. 4.1 to 4.4. As can be seen in all figures, both temperature and pressure fluctuate during calculation time. The fluctuation is observed to indicate the real physical features normally occurred in the canonical system; here, the temperature is not a well defined variable even in the segregated system [85]. Also, the fluctuation is evidence since system's temperature is defined by thermostat temperature; here, the fluctuation is a common behavior in the interaction of heat-bath and the system's particle. In the case of pressure, this variable is generally the most thermodynamic condition to fluctuate. However, their average value shows constantly coincide with the desired value to achieve. During this production run, the feature is observed with an acceptable range of instantaneous value of fluctuation. Accordingly, it is safe to state that the calculation is under the condition similar to the experimental setup [33] and [34].

Meanwhile, average energy plot in this system shows constant value concerning the calculation time. Fluctuation is also evidence since the total energy and the temperature are related by [86–88] :

$$\langle \Delta E \Delta t \rangle = k_B T_o^2 \tag{4.1}$$

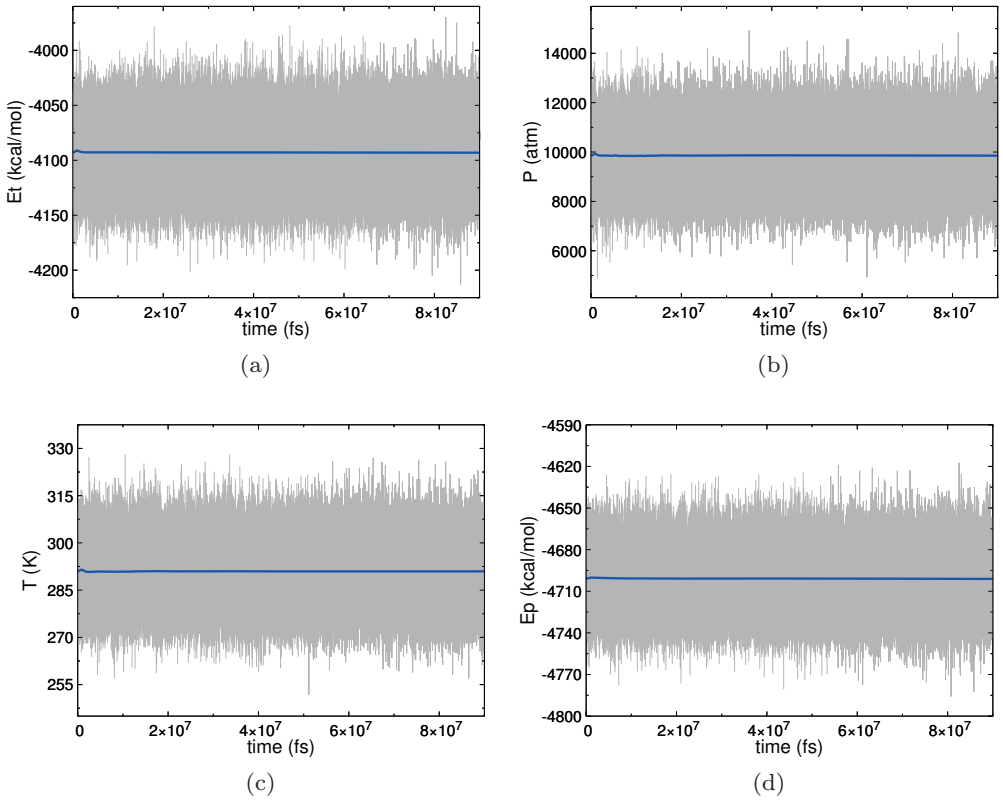


Fig. 4.1. Time series plot of Molecular dynamics calculation result on (a) total energy, (b) pressure, (c) temperature, (d) and potential energy C_1 full occupation system during 3 ns of production run. Grey line indicates instantaneous value while blue line for average.

where $\Delta E = E - E_0$ and E_0 is energy of system in equilibrium condition.

4.1.1 Filled ice II (C_1)

In hydrate C_1 system, average value of total and potential energy are depicted in Figs. 4.1a and 4.1d for full occupation, and Figs. 4.2a and 4.2a for the one vacancy. Plateau is shown on all plots of average energy value. These plots give evidence that the simulated system already in the equilibrium state and justifying the correct dynamic properties would have resulted from the production run. For the total energy, the constant value coexists at -4093(6) kcal/mol on full occupation, while -4093(4) kcal/mol coexist on one vacancy. Total potential energy is calculated to be -4700(2) kcal/mol for full occupation, while -4700(4) kcal/mol for one vacancy. The energy value is calculated from all molecular interactions to exist in the system. In the meantime, fluctuations are evidence showing regular physical feature acquired in most MD calculations. The total energy fluctuates between -4125 kcal/mol and -4000 kcal/mol for full occupation. Meanwhile, the total energy value varying from -4180 kcal/mol to -4030 kcal/mol in one vacancy setup. In the case of potential energy, fluctuation between -4750 kcal/ mol and -4640 kcal/mol observed in full occupation, while a similar range of values shown by one vacancy.

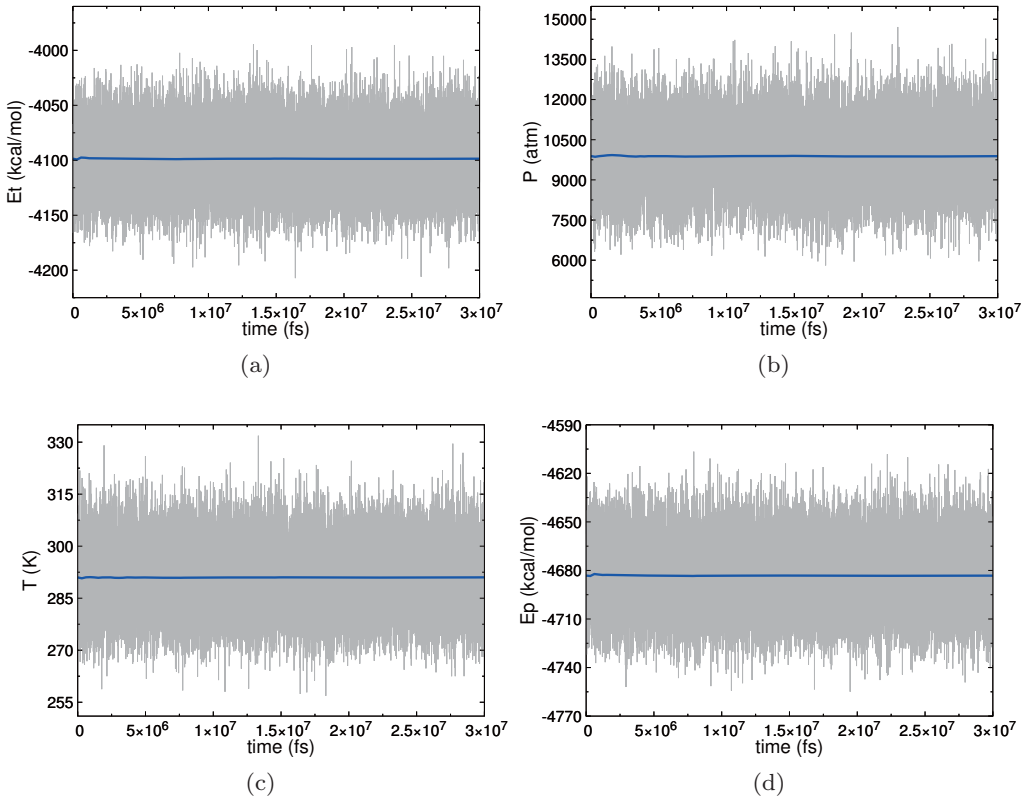


Fig. 4.2. Time series plot of Molecular dynamics calculation result on (a) total energy, (b) pressure, (c) temperature, (d) and potential energy of C_2 one vacancy system during 3 ns of production run. Grey line indicates instantaneous value while blue line for average.

The same plateau trend also noticeable in pressure and temperature plot. As can be seen in Figs. 4.1b and 4.1c for full occupation and Figs. 4.2b and 4.2c for one vacancy, the constant value are observed from the beginning of the production run and coincide with the desired value; here, the experimental setup in which the temperature of 291 K and pressure of 9869.23 atm are the desired values. As a consequence of applying thermostat and barostat, fluctuations are observed in these particular systems. The value varies in the range of allowable spectrum as occurred in common system studied by MD calculation. As expected, both distinctive lattice site occupation's arrangements show a similar range of fluctuations. The pressure fluctuates from 7000 atm to 14000 atm and the temperature varying from 260 K to 316 K. The same barostat and thermostat algorithm as applied in C_2 has given the desired value during this production run. From these features, it can be concluded that the all simulated system is already in the same thermodynamic condition as the experimental setup; here, fast diffusion of hydrogen guest molecules has been observed.

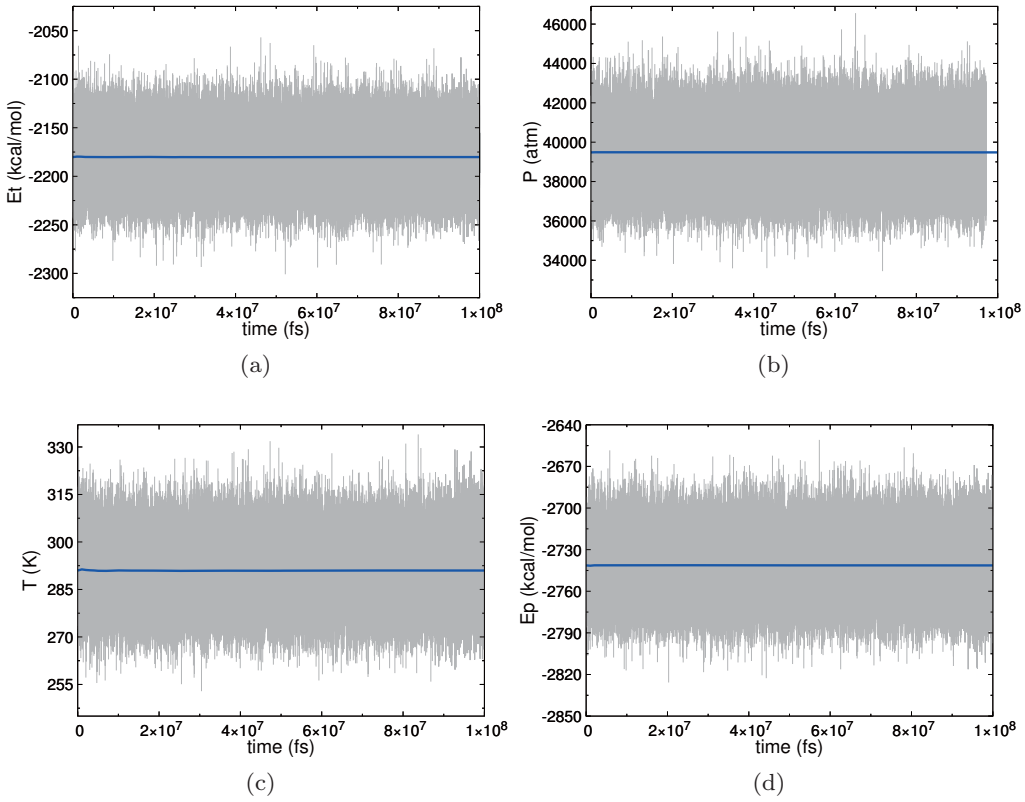


Fig. 4.3. Time series plot of (a) total energy, (b) pressure, (c) temperature, (d) and potential energy in full occupation of hydrate C_2 during 10 ns of production run. Grey line indicates instantaneous value while blue line for the average.

4.1.2 Filled ice Ic (C_2)

As can be seen from Figs. 4.3a and 4.4a, instantaneous total energy of the system are fluctuated between roughly -2250 kcal/mol to -2050 kcal/mol. These are observed in full occupation and one vacancy system. However, the average value of the total energy is constantly shown during the production run. It is calculated that -2180(4) kcal/mol is for full occupation and -2182(7) kcal/mol has resulted from one vacancy. Concerning the difference, these results show no significant distinct value since only one guest occupation peculiar between the two. It is also implied that no significant deformation or structural change would be occurred concerning system's structure. Correspondingly, potential energy calculation also shows the less contrasting value between the two setups. The calculated total-potential-energy average for 432 molecules is -2741(2) kcal/mol in full occupation. In the meantime, the instantaneous value fluctuates roughly between -2795 kcal/mol and -2675 kcal/mol. In the case of one vacancy, average potential energy is -2742(4) kcal/mol; here, the instantaneous value fluctuates between roughly -2790 kcal/mol and -2680 kcal/mol.

For the pressure and temperature, the thermostat and barostat are applied to the system's particle and heat bath through multiple chains. The scheme gave plateau-shaped at the beginning of production run. This constant average value coincides with the desired

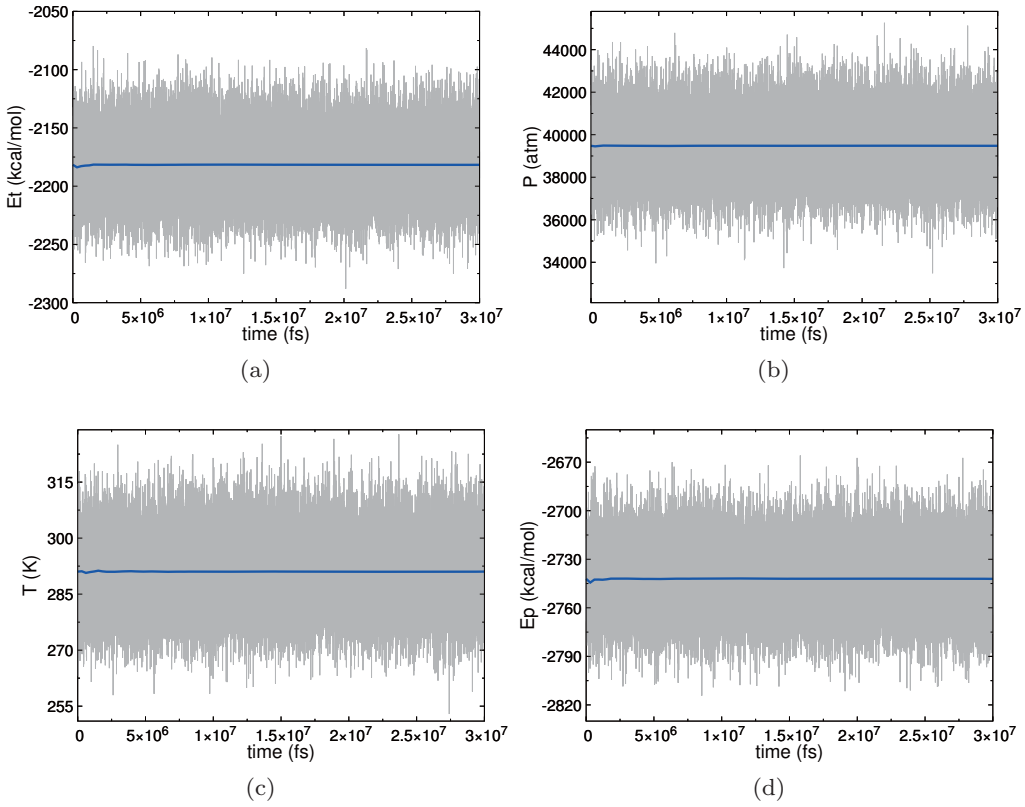


Fig. 4.4. Time series plot of (a) total energy, (b) pressure, (c) temperature, (d) and potential energy in one vacancy setup of hydrate C_2 during 1 ns of production run. Grey line indicates instantaneous value while blue line for average.

condition. In details, it can be seen from Figs. 4.3b and 4.4b, the desired value of 39476.93 atm was achieved during this production run. In the meantime, the instantaneous value fluctuation evidence between roughly 35000 atm to 44000 atm. In these hydrogen hydrate's systems, both two different setups of occupation show similar fluctuation's range. When considering system's temperature, the result shows a similar range of amplitude for both distinctive lattice site occupation setup. Similarly, the desired average value of 291 K was achieved in this production run. These results justify that the systems are equilibrated to give proper dynamical property under the desired experimental condition.

4.2 Energy and Density

Secondly, we compare system's energy and density between the one vacancy system and the fully occupied setup. Listed values are presented in Tab. 4.1 for system's energy per molecule, and Tab. 4.2 for the density. The comparison could be important to describe structural changes, if exist, in these two distinctive setups; here, the setups refer to full occupation and one vacancy. It can also be used to justify that an occupation-defected structure can also be found under the same thermodynamic condition as the ideally syn-

Table 4.1. Energy Comparison

System	Energy	Full Occupied (kJ/mol)	One vacancy (kJ/mol)
C_1	Potential	-52.03(2)	-52.16(4)
	Total	-45.31(4)	-45.43(7)
C_2	Potential	-26.55(2)	-26.62(4)
	Total	-21.12(4)	-21.18(7)

thesized. As can be seen from Tab. 4.1, both one vacancy system's have slightly lower values compared to the full. It is observed in all simulated filled ices structure. In the case of hydrate C_1 system, the different is only 0.13 kJ/mol to give 2.5×10^{-1} % contrasting on Potential energy and 2.7×10^{-1} % on the total energy. On the other hand, 0.07 kJ/-mol different is shown by hydrate C_2 system. The value showed 2.6×10^{-1} % deviation on Potential energy, while 2.8×10^{-1} % for the total energy. As expected, these small percentage would give no sign of structural change as has been observed. The lack of one guest in both simulated system only allowing hydrogen guest to diffuse without structural deformation.

When analyzing host framework's density as listed in Tab. 4.2, hydrate C_1 gives comparable value to the previous theoretical calculation [29]; here, the difference is only about 2 %. Meanwhile, about 10 % difference is shown by hydrate C_2. This deviation is due to the different model and thermodynamic condition used during the calculation. In C_2 system, previous theoretical result applied pressure of 2.1 GPa and temperature of 273 K [29], in which lower on both properties compared to the present work. Meanwhile, temperature and pressure of 300 K and 2.1 GPa, respectively were used in the previous calculation of hydrate C_1 [29]. As a result in C_1, the lower pressure used in this particular calculation gives lower host density value. In the case of C_2, as expected, higher water density value is calculated due to higher pressure applied in the present system.

Furthermore, concerning system's density, the comparison will be mainly emphasized on lattice site occupation setup. In the case of hydrogen hydrate C_1, a higher value on both total and water density are observed in one vacancy. The different is about 8×10^{-2} % in water density while 4×10^{-2} % in the total. In the case of C_2 system, a similar total density of 1.192 g cm^{-3} is shown for both setups. On the other hand, 0.001 g cm^{-3} higher water density is shown by one vacancy. This contrast is expected since only one hydrogen molecule differentiating between the two. These results can be explained as follows. Since controlled high-pressure condition was applied when performing the calculation, all of the one vacancy setups receded their volume. The absence of a repulsive interaction's contribution from a guest molecule in a lattice site is believed to be the origin of this feature. However, due to the higher stoichiometry of hydrogen and water molecules in C_2 system, the lack of one guest molecules showed less effect to the overall system's density.

The previous study that is used as a comparison in this study reveals discontinuity in water density value due to the structural collapse in C_1 upon decompressing [29].

Table 4.2. System's Density

System	C_1 (g/cm^{-3})		C_2 (g/cm^{-3})	
	total	water	total	water
Full Occupation	1.192	1.170	0.9577	0.8616
	Ref. [29]	1.195		0.775
One vacancy system	1.192	1.171	0.9581	0.8623

Additionally, since hydrogen gas in high temperature and pressure showed higher density compared to its condensed phase at a lower temperature, more dynamics is expected to be observed; in a liquid nitrogen temperature, the thermal motion is still substantially featured by hydrogen molecule [34].

4.3 Diffusion Mechanism

Thirdly, diffusion process occurred in full occupation setup is discussed. The structure was prepared to demonstrate the ideally synthesized filled ices. All lattice sites were accommodating one hydrogen molecules. Initially, those guests were positioned in lattice sites to eventually aligned to an energetically favorable position during the minimization stage. Lattice site volume in hydrate C_1 has more significant value compared to the C_2. During 100 ns of calculation time, each hydrogen guest molecule oscillates within the lattice sites. Eventually, no inter-site translation performed by hydrogen guest molecule in hydrate C_1 and C_2. Oscillation is found to be the only movement observed in these full occupation setups.

Next, the diffusion process is discussed in one vacancy setup. This arrangement has one less hydrogen molecule inside its ice structure compared to the full setup. One lattice site is left empty to show the imperfect synthesized result. In contrast with the full, oscillation and inter-site movements are evidence during 50 ns of calculation time. Many jumping processes are observed to indicate rapid diffusion in this structural system for hydrate C_1 and C_2. To further confirmed the finding, the diffusion coefficient of hydrogen guest molecules is calculated on both hydrates. The calculation is performed by using Eq. [3.4] and employed with the following stages. Firstly, the mean square displacement (MSD) of tagged hydrogen guest molecule is computed by using:

$$\langle |\Delta \mathbf{r}(t)|^2 \rangle = \langle (\mathbf{r}(t) - \mathbf{r}(0))^2 \rangle \tag{4.2}$$

where $\mathbf{r}$ is tagged hydrogen guest position vector and $\mathbf{r}(t) - \mathbf{r}(0)$ is displacement vector that has been traveled for t period of time. The plot is presented in Fig. 4.8. It can be seen from the figure that at the shorter time correlation, both systems show hydrogen molecule oscillates inside the lattice site. This movement is indicated by a very steep inclined line in MSD graph. The diffusion coefficient is then calculated by using Eq [3.4] in which the tangent value is derived from MSD's curve fitting. The results are then listed

Table 4.3. Diffusion coefficient of hydrogen molecules (D) and vacancy (D_{vac}) in units of cm^2s^{-1}. C_1 and C_2 are for filled ice II and Ic, respectively.

System	D	D_{vac}	Exp.
C_1	3.9×10^{-8}	8.0×10^{-7}	1.3×10^{-8} [33]
C_2	5.2×10^{-9}	1.8×10^{-6}	$\simeq 10^{-9}$ [34]

in Tab. 4.3.

The experimental study which reported the rapid hydrogen molecules diffusion to be comparable to the molecular to diffuse in the liquid state is used as a comparison. These studies are under the same thermodynamic conditions as this present calculation set up. Previously, there was two diffusion coefficient to compare with this particular hydrate C_2 system. The diffusion coefficient previously showed in [33] was about 5×10^{-11} cm^2 s^{-1}. This value have resulted with less extensive analyze on hydrate C_2 system and further revised in [34]. From this experimental results, fast diffusion of hydrogen molecules is proven as a universal feature observed in filled ices, showing less effect on the highly applied pressure. The new higher value then agrees more with the present calculation as well in hydrate C_1, confirming the active translation of hydrogen molecules among lattice sites. This means that the same rapid diffusion is observed in this MD calculation. Also, the agreement give suggestion that the real physical phenomena is captured in the calculation. In conclusion, these justifies the provided detail diffusion mechanism presented in this research.

Vacancy diffusion coefficient is also calculated in these filled ice systems. The calculation was performed simultaneously with the identification of inter-site hopping process. An array of oxygen atoms within a specific distance range from the hydrogen molecules inside a lattice site was screened and tabulated; here, oxygen atoms of water molecules are coordinated on the vertices of polyhedral lattice site. In doing this particular screening, filled ice configuration where all of the lattice sites accommodating one hydrogen molecule was used. In hydrate C_1 system, distance of 4.2 Å from hydrogen molecule hooked 12 neighboring oxygen atoms assembling a polyhedral lattice site. In the case of hydrate C_2, 4.3Å was used to list 10 oxygen atoms position in which located at the vertices of polyhedral lattice site. Subsequently, the lattice site's midpoint was calculated by using the listed oxygen atoms obtained from one vacancy system trajectory. Next, every hydrogen molecules at a time were labeled with their accommodating lattice site. During the simulation time, inter-site hopping was detected by the change of lattice site label addressed on hydrogen molecules. The vacancy position was then depicted from the lattice site's midpoint where the identity was previously assigned to the hopping hydrogen molecule. The time series of this lattice site's midpoint was later used in the calculation of mean square displacement and vacancy diffusion coefficient. In the hydrate C_1, the method give the diffusion coefficient of vacancy 8.0×10^{-7} cm^2 s^{-1}. On the other hand, the unoccupied lattice site diffused with the diffusion coefficient of 1.8×10^{-6} cm^2 s^{-1} in the hydrate C_2. The unoccupied lattice site in both systems show faster diffusion compare

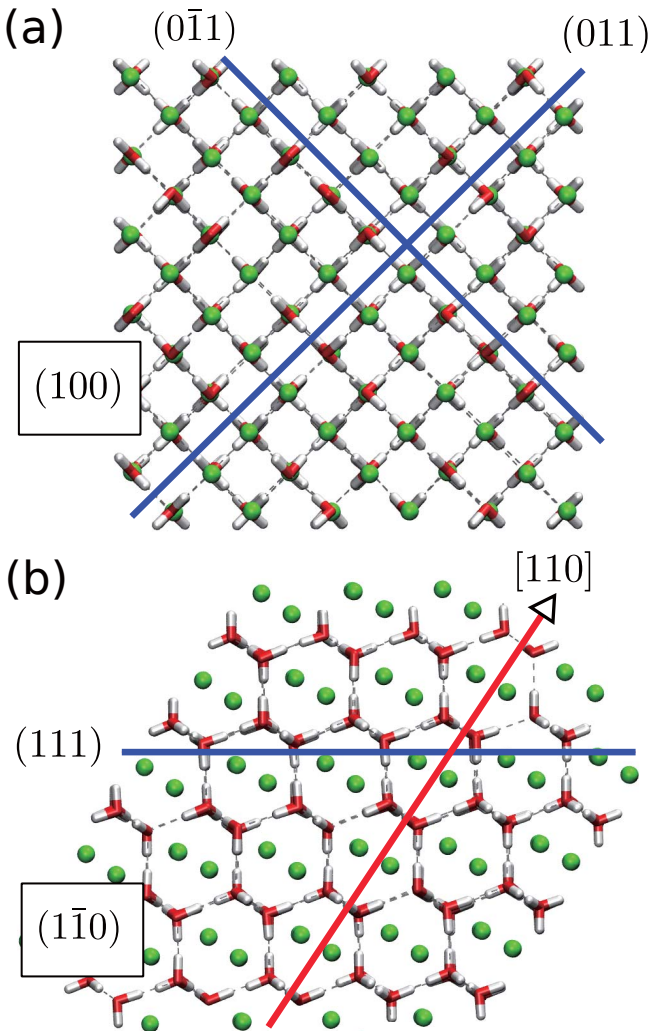


Fig. 4.5. The arrangement of water and hydrogen molecules in hydrate C_2. Water molecules are drawn by the red-white licorice shapes and H_2 molecules described by the green spheres. Hydrogen bond connecting the water molecules is shown by dashed gray line. Text on the box explains the drawing plane of the framework; here, the system is drawn on (a) (100) and (b) (1$\bar{1}$0) plane. Blue lines crossing the structure and the text next to the line show the crystallographic surfaces in which characterize the system. Red arrow is pointing direction of [110] as illustrate in the figure.

to the hydrogen molecules; this trend is more remarkable for the C_2.

In the case of C_1 system, guest molecules can only do inter-site movements in the direction of crystallographic c axis in hydrate C_1, with the average separation distance is 3.23 Å. As shown in Fig. 3.2 (a), the highly anisotropic diffusion can be explained by the neighboring separation along basal plane, in which almost three times larger than the separation along optical c axis. Many hopping effects performed by hydrogen molecules observed in this hydrate. There were 383 hydrogen guest inter-site translation for 50 ns of the calculation run with the first occurred at 3.6619 ns.

In contrast, as can be seen in Fig. 3.2 (b) and further detailed in Fig. 4.5 for hydrate C_2,

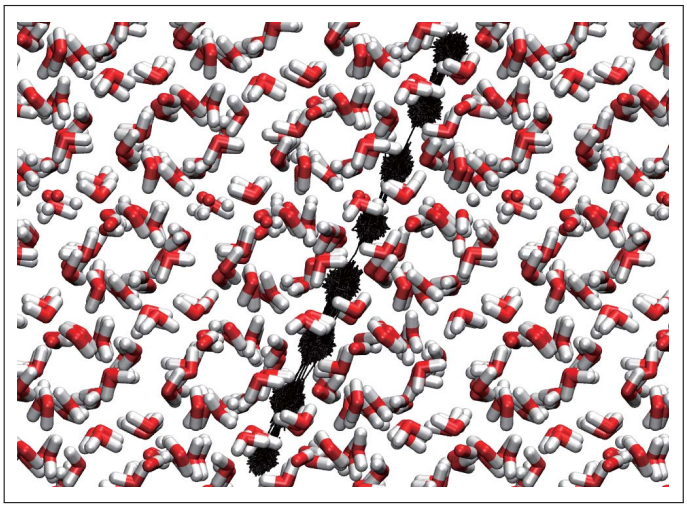
guest molecules are arranged such that "bond" connecting the adjacent 4 guest molecules are coordinated parallel to the tetrahedrally arranged of water molecules. The bond is pointing towards directions normal and parallel concerning plane $(0\bar{1}1)$ of cubic ice structure. Uniform distances between lattice sites along the line are observed in this hydrogen hydrate system. These arrangements suggested that isotropic guest diffusion can be found in this hydrate. Indeed, during the production run, many inter-site translations that are investigated follow the pathway coincides with these planes. The average guest distance is calculated to be 2.78 Å. There are total 819 of guest hopping, where 393 evidence on the plane parallel to the cubic ice $(0\bar{1}1)$ plane while 426 occurred along the (011) surface.

Next, the diffusion process of hydrogen molecules is discussed. In doing this, a hydrogen molecule trajectories are depicted from one vacancy system. The hydrogen molecule in which have the most frequent inter-site movement is depicted to demonstrate the diffusion process. The evolution of this tagged molecule can describe the general way of guest species when performing the hopping. As can be seen in Fig. 4.6 for one vacancy systems, during the production run, most movements observed are the oscillation in its lattice site. Jumping is found involving two neighboring sites; here, the inter-site change only evidence if the adjacent polyhedral lattice site is unoccupied. These inter-site movements are noticeable for both filled ice systems of one vacancy. However, hopping trajectories are found to depend on the crystallographic planes of filled ices.

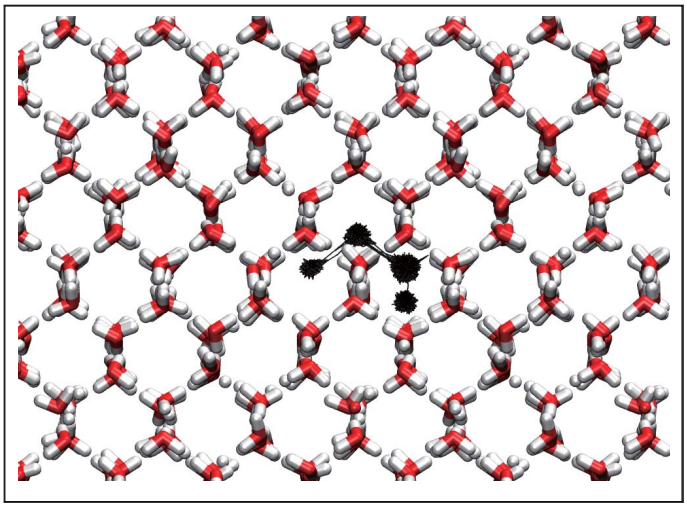
It is hard to predict which hydrogen guest molecule will do hopping in the next consecutive time frame. In other words, it is effort full to determine whether the next inter-site translation is the turning back or the other guest hopping. This difficulty evidence on both simulated hydrate systems. However, care-full observations were done to indicate the jumping process. It is found that the jump only involving one guest at a time with no correlated hopping evidence. The snapshot of these movements is shown in Fig. 4.7. Evolution of one tagged hydrogen guest molecule on each filled ice system is used to demonstrate the inter-site movement.

In the case of C_1, details of inter-site movement is drawn in Fig. 4.7b. Here, the tagged hydrogen guest molecule where initially oscillate in its previous lattice site performed inter-site translation at 110.8 ps. The path to follow is parallel to the optical c axis. Represented by yellow spheres, two stages are observed in this hopping. Distance covered by the first stage of jumping is 2.55 Å. Meanwhile, 1.49 Å is covered from the next stage of 100 fs. After the jumping process completed, the calculation revealed that the previous occupied lattice site is then engaged by the other hydrogen molecule. This occupation thus prohibiting the guest occupying its previous vacated lattice site in the next trial jump. Oscillation within the entire lattice site space is the only noticeable movement for this hydrogen guest during the roughly 1.6 ns.

In the case of hydrate C_2 shown in Fig. 4.7a, there are three drawn inter-site translation. The sequence are represented by yellow spheres as the first, green as the second and the last is shown by blue spheres. In an earlier time-frame (yellow spheres), the tagged hydrogen guest performed inter-site translation parallel to the $(0\bar{1}1)$ plane. The jumping



(a)



(b)

Fig. 4.6. Trajectory of the most frequent guest molecules to do inter-site translation in (a) C_1 (b) C_2 of one vacancy system. Water molecule's framework are shown by the red-white licorice shapes. Black solid line represents the path taken by a tagged guest molecules during 50 ns of calculation time. Hydrate C_1 is viewed along $(1\bar{2}1)$ of ice II (rhombohedral unit cell with spacegroup $R\bar{3}$) and C_2 is drawn on (110) plane.

took 200 fs to complete its two stages. In the first stage where 100 fs is taken, the guest is still in the previous void to left abandoned in the next 100 fs of second stage. This jumping scheme happened at 24.875 ns. Subsequently, during the next 12 ns roughly, the guest only undertaking oscillation movement inside its new lattice site position. While oscillating, trial turning-back jump is probable to be performed by the tagged guest in the meantime. However, the vacated lattice site is then occupied by other guest molecules, thus prohibiting the turning- back hop. The next inter-site movement occurred at 36.2858 ns. This particular translation fulfilling the empty void next to the present within distance parallel to (011) plane. Similarly, the two stages are needed to complete the hop in 200 fs. Slightly different in the covered distance shown by this second inter-site movement

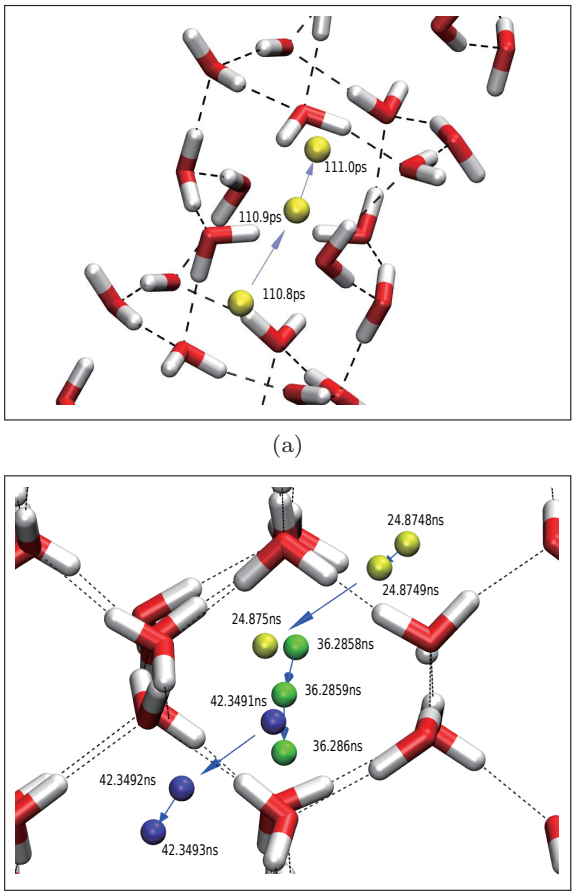
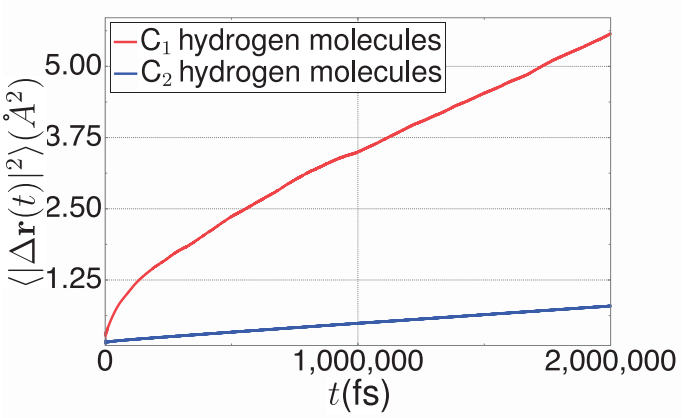


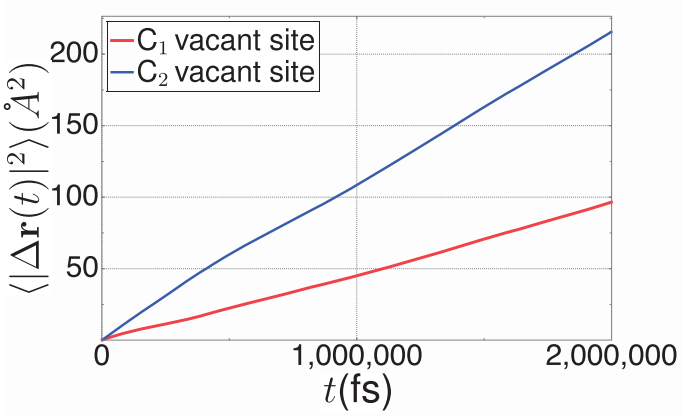
Fig. 4.7. Snapshot of a guest molecule performing inter-site hopping in orthographic view for one vacancy system of (a) C_1 and (b) C_2. Dashed black lines indicate hydrogen bonds while blue arrows show the direction. Colored spheres (green, blue and yellow) are for guest molecules. In both figures, spheres with similar color are used to represent the same hopping scheme in successive time. In the case of C_1 system, both yellow spheres indicate the guest movement along the crystallographic c axis. inter-site hydrogen guest trajectory parallel to the $(0\bar{1}1)$ plane shown by yellow and blue spheres in C_2 system, while the movement to the (011) plane is represented by green spheres.

compared to the first. During 100 fs of the early stage, the tagged molecules already in its halfway towards the new lattice site. After the jumping completion in the next 100 fs, the tagged guest oscillates in its new lattice site. It is probable to make trials jump to the previous lattice site. However, the empty vacated lattice site was then occupied by other neighboring guest. The third inter-site translation is observed at 42.3491 ns. Shown by blue spheres, the diffusion followed the pathway parallel to the $(0\bar{1}1)$ plane. During the first stage of 100 fs, the tagged molecule already moved to the new lattice site. Oscillation movement is then noticeable during the next stage of 100 fs.

Furthermore, potential field is calculated for all of the one vacancy systems to provide more evidence on the diffusion mechanism. Adopted from Ikeda-Fukuzawa et al. [89], the calculation was done through the following steps. Firstly, a tagged molecule and its



(a)



(b)

Fig. 4.8. Mean square displacement (MSD) as a function of time t for (a) hydrogen molecules in one vacancy system of filled ice II and Ic, and (b) vacant site of hydrate C_1 and C_2.

time-frame right before the hopping performed were chosen. Secondly, the planes in which parallel and normal concerning the crystallographic surfaces were prepared. Thirdly, these planes were then gridded with equal spacing of 0.1 Å. Tagged hydrogen guest was then moved on to the gridded point to calculate the total potential energy of this new configurations subsequently. Finally, the computed values were then contoured with interval equation is defined by

$$E_n = a \times 4^n \,, \tag{4.3}$$

where a is the starting value of potential energy, and n is the number of isolines. The results are drawn in Figs. 4.9 and 4.10.

As can be seen in Figs. 4.9 and 4.10, Many potential barriers observed to encapsulate the occupied lattice site. These inclusions thus prohibiting the tagged molecule to fill the already accommodated lattice site. In the same figure also noticeable a relatively flat energy area enclosing tagged guest position (black triangle). This lowest fields area is spanned to cover spaces where the vacancy is located. To get a clear description regarding

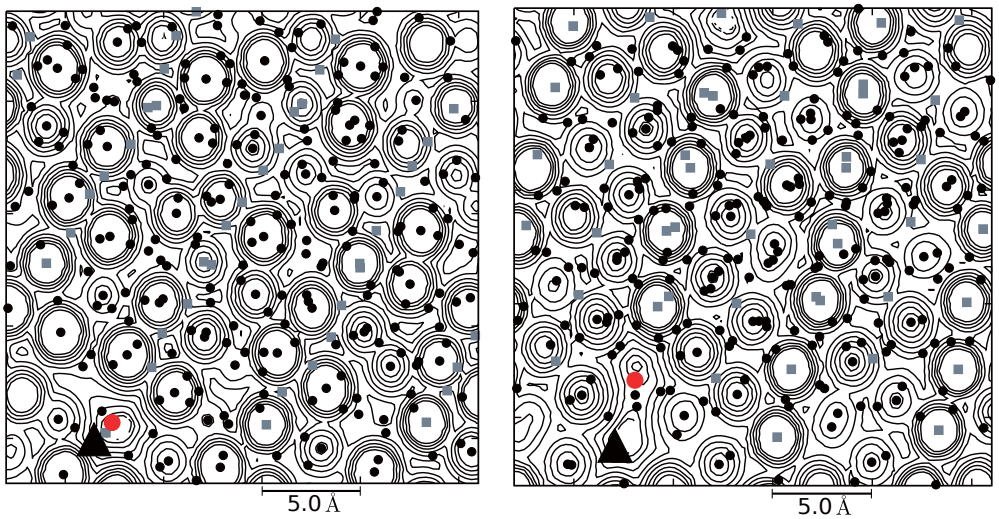
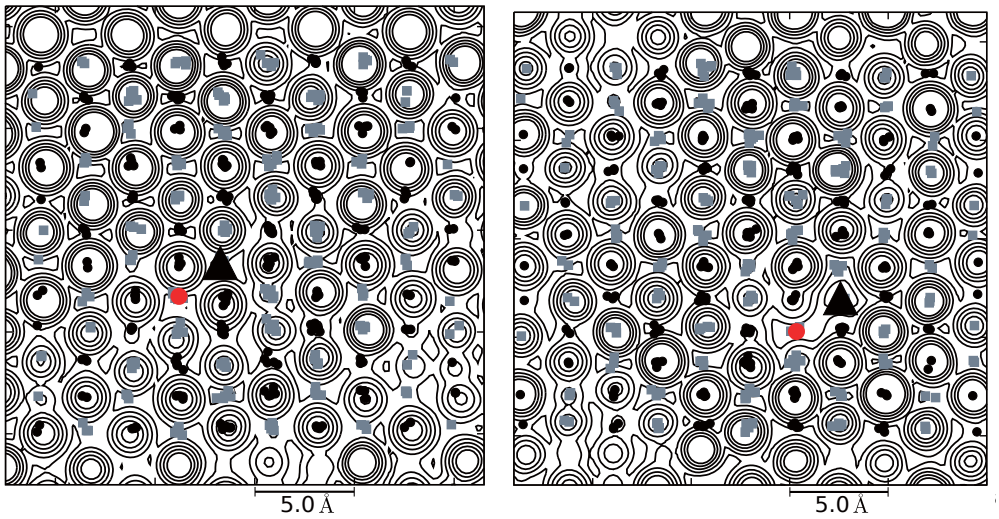


Fig. 4.9. Potential Energy contour of one vacancy C_1 system, snapshots are drawn at 0.36619 ns on projection surface parallel to (a) basal and (b) optical plane. On the basal plane, 10 contour lines are used with the lowest energy to describe is 1 kcal/mol and 2.6×10^5 kcal/mol for the highest. On the other hand, 14 contour lines are used on the optical plane in which the lowest energy to represent is 6.25 $\times 10^{-3}$ kcal/mol. Hydrogen guest and oxygen-water molecules that are projected on the plane are represented by gray rectangles and black spheres, respectively. Black triangle as shown in C_2 system is also used to describe the hydrogen guest molecule to have inter-site movement in the next consecutive time-frame. Red sphere is drawn to denote the vacancy location. As previously shown by the C_2 system, the tagged hydrogen guest to jump in this C_1 system is located in the area having the lowest potential energy value. In the case of energy map on optical plane, the lowest energy area spanned wider surfaces to cover empty lattice site to occupy in the next consecutive time-frame. In contrast, the lowest energy area on the basal plane is shown very small to cover both the position of tagged hydrogen molecule and the vacancy. Energy barrier built around the triangle restricted the diffusion process to occur on this basal plane.

inter-site movement, pathway followed by this hydrogen guest in the next consecutive time frame is evaluated. The path is found on the same area where the tagged guest and empty lattice site are positioned. However, the pattern is not evidenced on basal planes potential field of hydrate C_1. As can be seen in Fig. 4.9a, the inclusion covers an area to distance less than the tagged-guest and empty lattice site separation. There are boundaries between the tagged molecule and the nearest vacancy along this basal plane. Meanwhile, as shown in Fig. 4.9b, the pattern indicates inter-site movement evidence on a plane parallel to the crystallographic c axis. These features thus further confirm the highly anisotropic diffusion labeled in hydrate C_1 systems. Meanwhile, no energy barrier can be clearly seen in Figs. 4.10a and 4.10b for hydrate C_2. This energy barrier pattern justifies that the inter-site movement is isotropic in the direction concerning plane (110) and $(\bar{1}10)$ in hydrate C_2.



(a) (b)

Fig. 4.10. Contour plot of one vacancy hydrogen hydrate C_2 potential energy drawn on surface parallel to (a) hexagonal (110) plane at 0.40106 ns and (b) ($\bar{1}$10) plane at 0.60572 ns. There are 8 isolines to describe potential energy map for the hexagonal puckered-surface, with the lowest value describing by the line having potential energy of 64 kcal/mol. Similar to the previous plane, eight increasing pair potential energy lines are used to describe the map on the normal. The plotted energy line represent energy value started from 16 kcal/mol. In both figures, guest molecule to do inter-site movement in the next time-frame is represented by a black triangle while the vacancy is denoted by red sphere. The lowest energy area where the hopping candidate is located to cover spaces where the vacancy is settled. Black spheres and gray rectangles denote the projected oxygen-water and other hydrogen guest molecules on the plane, respectively.

4.4 Radial Distribution Function

This section will discuss radial distribution function (RDF) of both simulated systems. The discussion mainly about the comparison between distinctive lattice sites occupation encompassed by the same filled ice system. The structural difference between these two hydrate systems also discussed. RDF profiles of all simulated system are presented in Figs. 4.11 and 4.12, to be presented as a starting point of discussion. Comparison between the distinctive accommodation setup is performed to determine whether there is structural change evidence during calculation. This study could be physically relevant to justify the existence of occupation defect in filled ices. All atomic pair RDF, which is part of the system, are calculated. RDF profile of O^w-O^w and H^w-H^w indicates both systems still have the basic structure of their respective proton ordered of Ice II and disordered of ice II structures [90]. However, it is expected to have slightly different when the ices structure is accommodating guest molecules inside their lattice site. The difference is mainly on the peaks location, width, and magnitude while keeping the general pattern's composition. To start with, the first peak of O^{water} - O^{water} pair corresponds to the distance of two point edges of hexagonal rings in host water framework. The presence of guest molecules

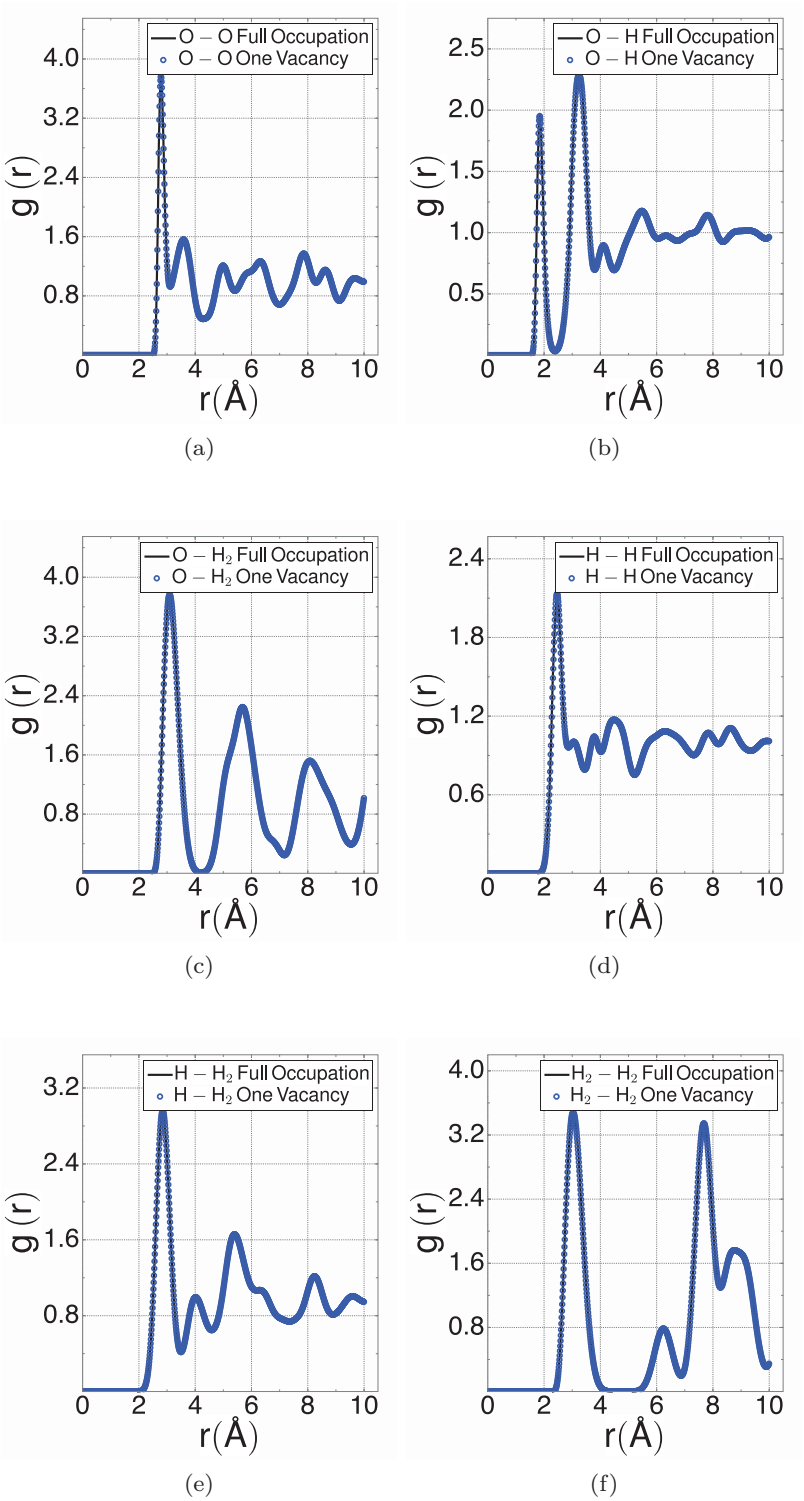


Fig. 4.11. Radial distribution function plot of (a) O^w-O^w, (b) O^w-H^w, (c) O^w-H_2^g (d) H^w-H^w, (e) H^w-H_2^g, and (f) H_2^g-H_2^g pairs in C_1 system.

have made the RDF of O^w-O^w pair of all simulated system peaked at around 2.8 Å. The distance is a typical behavior found in bulk-like water system interacting with the

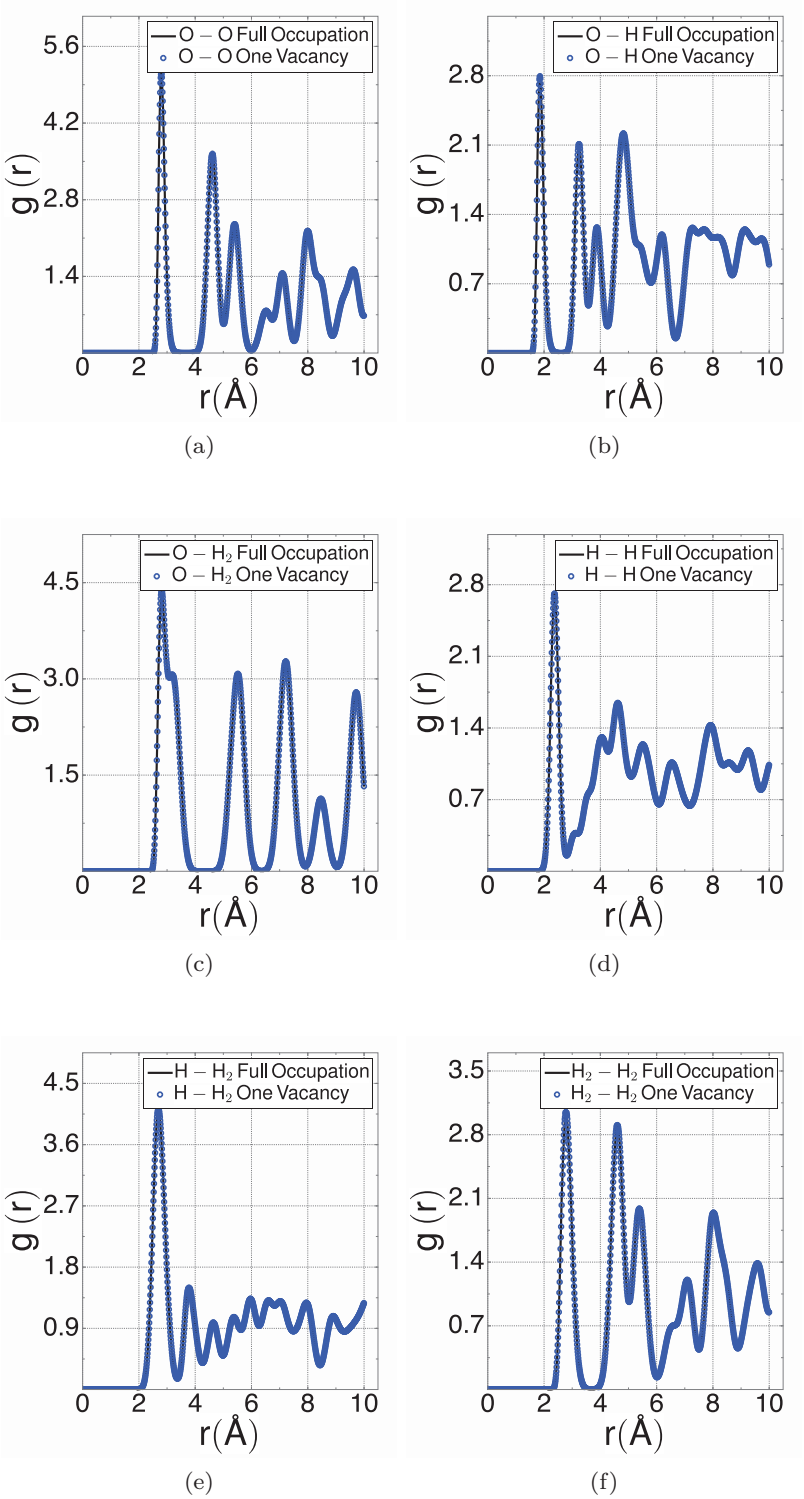


Fig. 4.12. Radial distribution function plot of (a) O^w-O^w, (b) O^w-H^w, (c) O^w-H_2^g, (d) H^w-H^w, (e) H^w-H_2^g, and (f) H_2^g-H_2^g pairs in C_2 system.

hydrophobic molecule such as hydrogen.

The structural differences between those two hydrates are further shown starting from

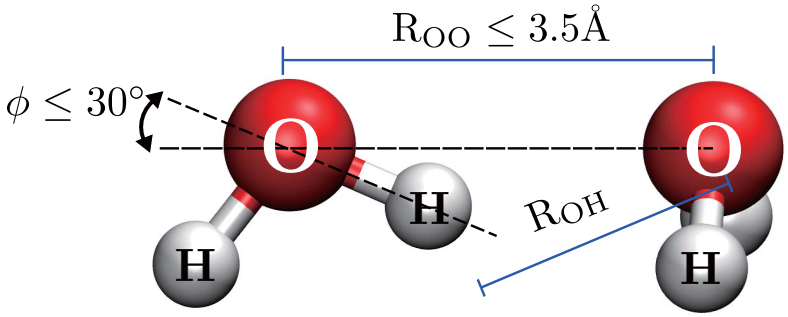


Fig. 4.13. Geometry of hydrogen bonding criteria used in the calculation

the second peak. Hydrate C_1 have this second peak at 3.5 Å in which showing "hump-like" shape. Meanwhile, the second peak with low magnitude is located at 4.635 Å in the case of C_2. This peak positioned at a pair-atomic distance comparable to the third peak of hydrate C_1. This feature is noticeable since a unit cell of C_1 consist of two hexagonal rings closely approaching each other. The rings separated within distance shown by the second peak with no hydrogen bonding connection. The next spike further indicates structural differences between these two. In principle, the distinctive evident concerning the way of hexagonal ring's edges and the arrangements of its neighboring.

Concerning the lattice site loading preferences, full occupation and one vacancy show no significant differences on their RDF plot. The absence of one repulsive character governed by host-guest interaction do not change the structure. These are found on hydrate C_1 and C_2. The defect is only facilitating hydrogen guest experiencing rapid diffusion under these high thermodynamic conditions.

The discussion is then performed on RDF of H_2^g-H_2^g pair. It is observed that the first peak in hydrate C_2 appeared at 2.795 Å, which is found to be 3.03 Å in hydrate C_1. This distinctive distance is due to the lattice site volume contrast found between these two hydrate systems; here, C_1 have bigger lattice site volume than C_1. The similar differences can be found in the first peak of RDF O^w-H_2^g between these two hydrates to further clarify the different lattice site volume. Meanwhile, the magnitude of the second peak of RDF H_2^g-H_2^g pair is found to be similar to the first in hydrate C_2 system. This pattern describes the uniformly neighboring distance concerning the crystallographic planes as shown in ice Ic structure. In contrast, as observed in hydrate C_1 system, the second peak's magnitude found to have a huge difference compared to the first. This magnitude contrasting rationalized the guest neighboring distance three times farther along basal plane compared to the optical c axis. The peaks give n_{th} distance order concerning the optical axis.

The isotropic and anisotropic term difference is further explained by the number of peaks found in O^w-H_2^g pair. There are five peaks appeared in hydrate C_2 while three wider peaks are shown in hydrate C_1 system. The freely moving guest molecules to oscillate in hydrate C_1 lattice site have made wider distance's probability in finding the n_{th} nearest order of guest molecules concerning the O^w position.

Analysis of O^w-H_2^g and H^w-H_2^g give guest molecules are favorable to have a shorter distance to the hydrogen water. Tail (as shown in ref. [91]) is not evidence in the interaction distance below 2.5 Å of H^w-H_2^g since the respective system to study have different ice framework.

In this RDF profile comparison, the first peak appeared at a slightly shorter distance in C_1 system. This distance corresponds to the hydrogen bonding regime in the framework. Along with O^w-O^w pair, both give the larger bond energy calculated from C_1 system as can be seen in table 4.4.

The pair of O^w - H^w RDF of all systems are explained with the additional calculation of hydrogen bond energy and distance. With regards to the structure, hydrogen bond distance and interaction energy by considering the geometry recognition proposed by Luzar-Chandler [92] are observed. The geometry criteria of hydrogen-bond distances and energies that are used in the calculation are depicted in figure 4.13.

Furthermore, hydrogen bond distance and interaction energy are also calculated. The bond was identified by using geometry criteria proposed by Luzar and Chandler [92]. At this temperature condition, hydrogen bond network of C_2 has been suggested to be disordered, as well as the orientation of hydrogen molecules [23, 93]. Meanwhile, the ordered hydrogen bond is observed in C_1 system. The calculated value of hydrogen bond distance and energy are listed in Tab. 4.4. The average hydrogen-bond distance in C_2 is 2.812(4) Å and the energy magnitude is 5.64(4) kcal/mol. On the other hand, the calculation on C_1 give 2.804(4) Å and 5.81(3) kcal/mol for hydrogen-bond distance and energy, respectively.

Furthermore, the energy and distance values are compared to the values of other stable hydrate systems, namely SI and SII, to analyze the physical properties that are contained in the resulting values. These two referenced clathrate hydrates are known to be the most stable hydrate structure in most of the thermodynamic conditions [55]. The comparison also performed with the result of unoccupied hexagonal ice. It has been reported previously that SI and SII hydrate have lower thermal conductivity compared to ice Ih [94]. Chakraborty et al. [95] showed that hydrogen bond energy of SI structure is the highest compared to SII and ice Ih. These results justifying that the more strained hydrogen bond to limit phonon propagation is responsible for the lower thermal conductivity. Correspondingly, the resulted hydrogen bond energy of C_2 and C_1 shows only slight different value compared to SI and SII; here, C_2 shows the highest. These results confirm the structural stability of C_2 [93] and C_1 under this thermodynamic condition.

4.5 Infra Red (IR) Spectra

IR spectra calculation performed by Fourier transforming system's total dipole moment are presented in Figs. 4.14a and 4.14b for both configuration setups of C_1 and C_2, respectively. Suppose that $\alpha(\omega)$ indicate absorption coefficient. Then, Fourier transform of $\boldsymbol{\mu}(t)$

Table 4.4. Hydrogen Bond Distance and Energy

System	r(Å)	E(kcal/mol)	Ref.
C_1	2.804(4)	5.81(3)	This work
C_2	2.812(4)	5.64(4)	
sI	2.775(4)	5.75(6)	
sII	2.815(1)	5.57(8)	[95]
Ice Ih	2.885(1)	4.72(7)	

time correlation function is given by [96] :

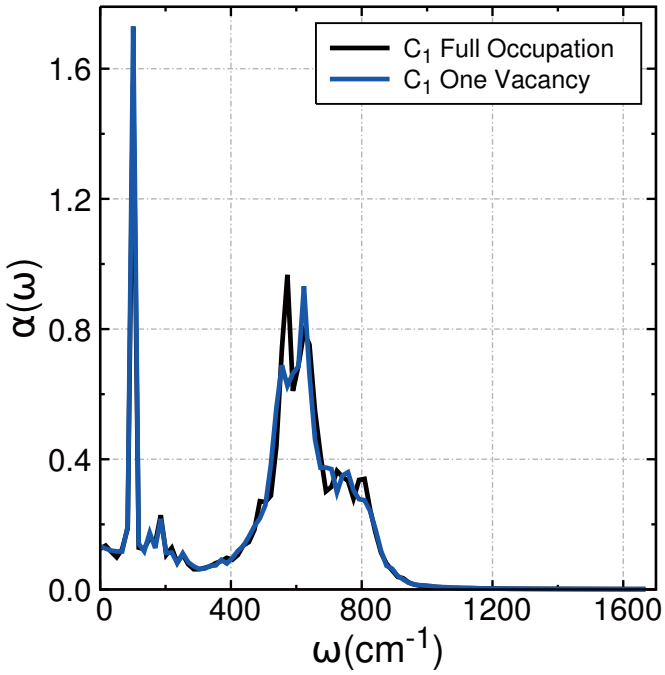
$$\alpha(\omega)n(\omega) = \frac{2\pi\omega\left(1-\mathrm{e}^{-\beta\hbar\omega}\right)}{3\hbar cV} \times \int_{-\infty}^{\infty} dt\, \langle \boldsymbol{\mu}(t)\cdot\boldsymbol{\mu}(0)\rangle \mathrm{e}^{i\omega t}\,, \tag{4.4}$$

where $\beta = 1/k_BT$, T is temperature, k is Boltzmann constant, c is speed of light in vacuum, V is the volume, $\hbar = h/2\pi$ which h is Planck' constant, and $n(\omega)$ is refractive index of the medium where the frequency is ω. In the case of a gas phase, $\boldsymbol{\mu}(t)$ is correspond to the dipole moment of the molecule. When calculating for fluid or solid, $\boldsymbol{\mu}(t)$ represents super cell's total dipole moment. In classical molecular dynamics, this $\boldsymbol{\mu}(t)$ can be generated by molecule's dipole moment according to the assigned force field.

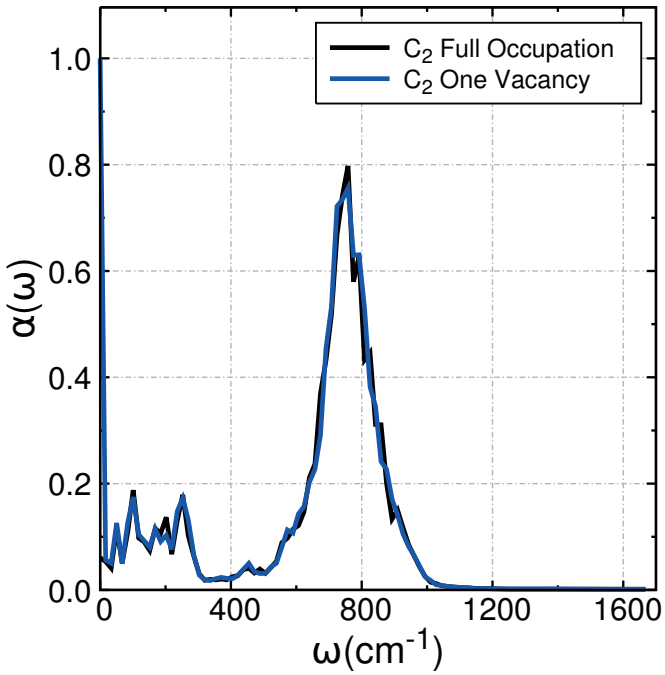
Framework's slight change of vibration modes when accommodating guest molecules can be studied by analyzing the respective unoccupied system's spectra. In general, The vibration data can be generated in several ways, including by employing inelastic neutron scattering (INS) of hydrogen bonding in ices as proposed in [97]. The experimental study produced at 15 K of temperature showed that structural transitions are the most probable to be observed compared to the combination and overtone. Besides, when analyzing the comparison, negligible hot band contributions were considered.

Concerning the rigid TIP4P/Ice water model when performing molecular dynamics simulation, low spectra calculation up to 1500 cm^{-1} is done. This low-frequency observation is presented since there will be no O-H bending and stretching spectra of water molecules observed within this rigid model. This bending and stretching occurred in high frequency above 1500 cm^{-1}. Both system's IR spectra plot shows water-librational motion, hydrogen-bond bending, and stretching spectra. The notable difference is on the absorption value. Neither frequencies of peaks nor the magnitude, comparison analysis between the full occupation and one vacancy give almost gives no difference. This no noticeable deviation of spectra could be originated from the usage of the one site model of hydrogen guest molecule and only one distinctive occupation setup.

In the case of hydrate C_1 system, six peaks in the dynamical HB region are observed. The same number of peaks can also be found in Ice II. The transverse acoustic mode has the highest magnitude compared to other spectra's peak located about 105 $\mathrm{c}^{-1}\mathrm{m}$. Meanwhile, the transverse optic mode peak which is describing the hydrogen-bond stretching can be found in roughly 184.16 cm^{-1} with less noticeable magnitude's height. The next spectra which are evident from approximately 370.32 cm^{-1} to 1006.16 cm^{-1} are also



(a)



(b)

Fig. 4.14. IR Spectra from total dipole moment correlation function of hydrate (a) C_1 and (b) C_2 system.

known as the L_1 and L_2 librational modes. The width of this transverse optic mode is higher compared to the empty ice II. It peaks at 592.60 cm^{-1}.

In hydrate C_2 systems, there are many peaks in the hydrogen bond (HB) stretching and bonding region of water spectra. The VDOS magnitude value of these hydrate's

frequency ranges is small compared to the librational mode. The same amount of peak with a slightly lower frequency also evidence in the ice I structure. Meanwhile, the next spectra, which is librational, is found in the range from 515.06 cm^{-1} to 1011.29 cm^{-1}. This spectrum peaked at 742.5 cm^{-1} while the unoccupied structure at 537.41 cm^{-1}. It is calculated that the librational spectra's width of the hydrate system is smaller compared to the empty ice structure.

Librational modes of C_1 system appeared slightly in the lower frequency with regards to the C_2, in contrast with the transverse acoustic and optic mode evidence at the higher. These results show a similar pattern when compare to their respective ice structure [97]. The higher energy of translational modes of C_1 indicate the less allowable guest molecules to do inter-cage in specific axis path, thus gaining repulsion interaction energy between host-cage since less permissible direction to diffuse.

To summarize the result, in current works, IR and Raman spectra are generated from MD total moment dipole data. Concerning the method, establishing a clear picture of intermolecular modes was slightly tricky. Correspondingly, the translational region is only represented by roughly ranged valued spectra as can be seen in table 4.5.

Considering this feature when performing the comparison to the original ice structure, the hydrogen hydrate system shows a slight blue shift in all spectra [97]. This deviation can be explained as follow. In the highly compressed order, hydrogen guest molecule is denser compared to its condensed phase. More dynamics are expected to be observed in the guest-hydrogen. The collision between hydrogen guest and its surrounding cage composed of water molecules framework give additional energy for the host to vibrate. The encounter also generates hydrogen guest molecular reorientation to modulate spin-rotation and dipolar coupling simultaneously [34].

Table 4.5. IR Spectra's peaks comparison

System	Spectra	This work(cm^{-1})	Ref. [97] (cm^{-1})
C_1	1^{st} Accoustic modes H-bond bending	105.72	100.82
	2^{nd} Accoustic modes	150.06	141.15
	3^{rd} Accoustic modes H-bond stretching	184.16	161.31
	4^{th} Accoustic modes H-bond stretching	218.26-255.78	191.56
	5^{th} Optic modes	266.94	241.97
	6^{th} Optic modes	324.85	302.46
	Libration band	370.32 - 1006.16	483.93 - 1109.01
	Libration's peak Major L_2 peak	592.60	504.10
C_2	1^{st} Accoustic modes H-bond bending	49.84	48.39
	2^{nd} Accoustic modes	105.72	96.78
	3^{rd} Optic modes H-bond stretching	167.70	157.28
	4^{th} Optic modes H-bond stretching	204-251.44	217.77
	5^{th} Optic modes Minor L_1 peak	302.46	453.03
	Libration band	515.06 - 1011.29	451.43 - 967.35
	Libration's peak Major L_2 peak	742.5	537.41

Chapter 5

Concluding Remarks

5.1 Summary

There are several clathrate hydrates where the host-water molecules framework is similar to ice structure. Instead of defining these structure as the clathrate, these hydrate arrangements which differ than the previously known clathrate structure described as filled ices. A similar non chemical-bonding interaction between guest and host-water molecules featured in this hydrate has given different phase boundaries and physical properties compared to the respective unoccupied ice structure. Several published data reveal that some of this compound was formed through clathrate structure pressurization.

Filled ice phases that have been discovered, namely C_0, C_1, C_2, MH-III, CO_2 hydrate-II and C_3 encapsulating guest molecules such as H_2 [23], CO_2 [61], CH_4 [64], He [30] and Ne [98]. Despite metastable state reported from several unoccupied ice structure, the occupied shows stable condition by means of single or multiple guest encapsulation.

In the case of hydrogen hydrate, fast diffusion of hydrogen guest in hydrate structure of filled ice revealed as main features which are interesting in the perspective of hydrogen storage material. This kinetic parameter enables one to get advantages of a faster synthesis process. Moreover, filled ice C_1 and C_2 which have maximum H_2: H_2O stoichiometry of 1:6 and 1:1, respectively, has been observed to have fast diffusion of guest molecules among cavities. The rapid inter-site translations are comparable to the molecular diffusion in a liquid state; here, this hydrogen-rich material compound in equivalent can contain a maximum of 10 % wt of H_2 capacity without having promoters to stabilize the host. Anisotropy and isotropy diffusion labeled for hydrogen hydrate C_1 and C_2, respectively, have given exciting feature on guest dynamics among enclosures. However, this behavior was left as a suggestion and has not been studied theoretically in the previous experimental research.

To provide evidence on diffusion process occurred in hydrogen hydrate filled ice II and Ic, molecular dynamics calculation has been employed under pressure condition of 1 GPa and 4 GPa, respectively; here, the temperature was controlled at 291 K for both systems. The chosen thermodynamic conditions are based on the fascinating-fast diffusion of hydrogen guest molecules found experimentally among lattice site formed by water

molecules framework.

There were many stages performed when doing the calculation. Firstly, the unoccupied structure of proton-ordered Ice II and proton-disordered of Ice Ic taken from Buch et al. [71] were prepared. Some single-site model of hydrogen guest molecules were then positioned into the center of every lattice site to have H_2: H_2O stoichiometry of 1:6 and 1:1 for C_1 and C_2, respectively. For the one vacancy setup, a guest in one lattice site was taken out from both filled ices system. The hollow lattice site was used to replicate the not ideally synthesized filled ices thus forming an occupation defect. Next, the energy minimization stage of these occupied structures was performed by using conjugate gradient method to subsequently employing the equilibration process for 50 ps. During equilibration, energy conservation and the fluctuation of controlled thermodynamic variables were carefully observed ensuring equilibrium condition was achieved under the desired temperature and pressure.

Oscillation with respect to lattice site's midpoint mostly spent by guest molecules during the calculation process. Repulsive characteristic observed in this guest-host interaction due to small lattice site is expected to show more dynamic phenomena, especially in these highly compressed systems. Under this thermodynamic condition, hydrogen molecules are denser compared to their condensed phase. Meanwhile, the longtime calculation is needed to study the diffusion process.

During 100 ns of calculation time, full occupation setup showed no guest inter-site translation for both C_1 and C_2 systems. Eventually, oscillation within lattice site only the noticeable movement performed by hydrogen molecules. The fluctuation was observed due to the momentum change of the guest when colliding the host with no next available vacancy. More significant space of oscillation is performed by hydrogen guest molecules inside hydrate C_1.

In contrast, it is found that the single unoccupied lattice site facilitates oscillation and many inter-site movements of hydrogen molecules for both systems during 50 ns of the production run. This feature suggested the necessity of an occupation defect to facilitate the inter-site movement in these filled ices. To further clarify the result, the Diffusion coefficient of guest hydrogen molecules is calculated. The result is in agreement with the experimental results; here, rapid diffusion of the hydrogen molecules was observed to be likely the molecular diffusion in a liquid state, and the result reproduces the synthesize observation. This agreement certifies that the same fast phenomena are found in this MD calculation. Also, the result justifies the detailed mechanism provided by this calculation to give the same real physical evidence as the experimental result has provided.

Furthermore, microscopic details of diffusion mechanism are discussed. During the production run on one vacancy setup, the inter-site movement is sequentially performed by the guest with no correlated jump evidence from the two hydrates. However, hopping trajectory followed by hydrogen guest molecules has revealed different character between the two hydrates. Host arrangements of C_1 system shows shorter distance along the crystallographic optical axis; here, closely tripled distance featured concerning the basal's separation is observed. This arrangements then rationalized the jumping can only be

found along the crystallographic *c* axis. With just following a pathway on one plane when performing inter-site jumping, highly anisotropic diffusion's term is then defining this hydrate C_1. In the case of C_2, the arrangements of host-water molecules compared to the guest showed closely similar and uniform in the distance. Also, the line connecting these molecules pointing toward directions in which coexist parallel to many crystallographic planes. The observations reveal that these directions are followed by hydrogen molecules when performing inter-site translation. When considering these features, isotropic diffusion is then labeled in hydrate C_2.

Diffusion uniformity can also be explained through energy map. The contour shows many energy barriers encapsulating the occupied lattice site. This arrangement was observed in hydrate C_1 and C_2. The model gives sign on the accommodated-lattice site's inability to be filled by the tagged and another neighboring guest in this setup. However, the barrier is not noticeable enclosing the tagged guest and the nearest empty lattice site. Both hydrogen molecules and vacant site are located in the same low-field area at a consecutive time before the jump is observed. The pattern is found on most of the studied crystallographic plane; the optical of hydrate C_1 and plane parallel and normal to (011) family in C_2 have this pattern. Meanwhile, the map on the basal plane of hydrate C_1 gives similar barriers on occupied lattice site with no low field area to cover the anticipated guest and empty lattice site. This particular pattern then explains the absence of inter-site movement on the basal plane of hydrate C_1.

The collision between guest and its enclosed lattice site structure increased host vibration energy. When comparing to the unoccupied system, this phenomena is evidence to make slight blue shifting on IR spectra calculation. The collision also generates hydrogen guest molecular reorientation to modulate spin-rotation and dipolar coupling simultaneously [34]. Meanwhile, the higher energy of translational modes of C_1 indicate the less allowable guest molecules to do inter-site in specific axis path, thus gaining repulsion interaction energy between host-guest. No noticeable difference between full occupation and one vacancy system found on both IR spectra calculation and Radial distribution function.

5.2 Future Work

Hydrogen hydrate C_2 has given research interest in the field of energy storage. In this material compound, hydrogen content can reach 10 % wt which is suitable for energy storage material. However, to synthesize this filled ice, the high-pressure condition is needed. Several published data revealed fast diffusion with less affected by pressurization is found under this highly compressed system. The material should be able to be generated under the more moderate thermodynamic condition to become available in a possible usage state. Also, it is the necessity that the product is stable for an extended time under ambient thermodynamic state concerning mobility and depository.

Kumar et al. [99] have synthesized hydrogen filled ice Ic from amorphous ice in low-

pressure condition less than 45 MPa. The process was performed by conditioning the amorphous ice in hydrogen gas pressure at low temperature. Subsequently, the annealing process to the higher temperature was implemented. The small amount of SII hydrate is then formed along with the crystalline structure of hydrate C_2 as the primary product.

Additionally, other fascinating points are the availability and stability at ambient pressure and low 77 K temperature. Despite less H_2 content to make this material under a possible condition to be the energy storage candidate, dynamic properties could be valuable to study. For instance, diffusion process of guest hydrogen molecules is associating with the time needed to complete the synthesize. Also, conversion kinetic gives more properties to be considered in the study.

Quantum effect of hydrogen guest dynamics is expected to be evident under this low temperature. It has been reported in hexagonal ice that quantum effect gives noticeable contrast in lattice vibrations and potential energy in comparison to the classically calculated counterpart [100]. Also, a molecular interaction governed by covalent bond of water when interacting with others shifted the phase-space boundary [101]. This effect also explained the inter-site hopping dynamics in clathrate hydrate structure dependable on temperature and lattice site accommodation number [102].

To further investigate the process, the long calculation is needed while revealing the quantum effect. Concerning the necessary features to be studied, Path Integral Molecular Dynamics calculation is suitable to be used. The method is implemented by using the isomorphism between the real quantum particle and the set of classical ring polymers. The formalism is as follows. The real quantum particle is represented by L replicas of particles in imaginary time sequences. These duplicates known as beads and connected to each other by a harmonic spring consecutively to form a ring polymer. The number of beads corresponds to the Trotter number while the ring to a cyclic quantum path. The classical equation of motion is designed in a such a way that the resulted partition function can be formally equivalent to the quantum set. The method is proven influential to study many-body quantum system.

A

Mathematical Backgrounds of MD

In this section, molecular dynamics simulation is briefly explained. Classical MD method is chosen to be used in these two filled ice system's calculation. Since the system under the thermodynamic condition where the quantum effect is negligibly small, the classical approach is sufficient to be used in the simulation of this condensed phase system. The molecular dynamics scheme is allowing one to proceed the extended period of simulation time with reasonable computational cost. Also, the method is proven to be statistically reliable in producing the required trajectory data for the study of a diffusion process.

LAMMPS molecular dynamics simulation package was used to simulate the system. This MD package offers an algorithm that can be used for large bulky system enhanced by the ability to be done in multi-platform parallel processes. The program can also deliver fast and accurate computation for extensive MD work. Within this dissertation document, several algorithms that LAMMPS closely followed are briefly derived and explained. The discussion is limited to only those that were used in this particular calculation. Subsequently, post "production run" analysis was done with additional self-build programs based on a commonly-used algorithm for the specific intention. The codes are only suitable for this particular purpose.

The derivation and discussion of the algorithm are started from periodic boundary condition scheme. The algorithm is suitable to be used in the case of orthogonal and non-orthogonal periodic cells calculation. In the next discussion section, the mathematical background of long-range energy contributions is explained historically, starting from the first proposed algorithm until the one that closely-followed by LAMMPS recently. Subsequently, the attached thermostat and barostat method are discussed. These algorithms are presented in historical sequences, in which similar to the previous algorithms discussion. Finally, the last discussion section in this mathematical background is the constraints algorithm used to incorporate rigid molecule model in the calculation. This final section will include a summary of SHAKE and RATTLE algorithm complemented with general stages concerning the implementation.

A.1 Periodic Boundary Condition

In post-production stage, the periodic boundary condition is calculated by using the following procedure. Suppose that the unit cell of the system is

$$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}} \tag{A.1}$$

$$\mathbf{b} = b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}} \tag{A.2}$$

$$\mathbf{c} = c_x\hat{\mathbf{i}} + c_y\hat{\mathbf{j}} + c_z\hat{\mathbf{k}}, \tag{A.3}$$

in which can be represented in the matrix form as

$$\mathbf{h} = \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{bmatrix}. \tag{A.4}$$

The particle's displacement vector is written by

$$d\mathbf{r} = dx\hat{\mathbf{i}} + dy\hat{\mathbf{j}} + dz\hat{\mathbf{k}}. \tag{A.5}$$

The periodic condition is applied by taking the matrix operation as follows. Let

$$\mathbf{h}_t = \mathbf{h}^t, \tag{A.6}$$

then the matrix of periodic boundary condition can be written as

$$\mathbf{B}_{\text{pbc}} = \mathbf{h}_\mathbf{t}^{-\mathbf{1}} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}. \tag{A.7}$$

By taking $\mathbf{B}_{\text{pbc}} = \lfloor \mathbf{B}_{\text{pbc}} \rfloor$ one can obtain the displacement vector subtracted from the periodic boundary condition as

$$\begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = \mathbf{h}\,\mathbf{B}_{\text{pbc}}. \tag{A.8}$$

A.2 Long Range Interactions

In this section, the long-range pair interactions are discussed. The discussion is constructed closely follow the historical sequence previously given by Frenkel and Smith [103]. The long-range calculation is important when one is trying to calculate a huge system. The computational cost would be greatly increased since great pair interaction should be incorporated. Therefore, the cost will be the main consideration in the biological system

where the calculated system may contain 10^5 particles to incorporate. Reducing the cost drastically by truncating the interaction distance to a certain radius can be done as previously been proposed. However, the calculation may be experienced lots of inaccuracies as a consequence. In the case of Coulombic and dipolar interactions, it has been proved that the truncation resulted diverge tail correction in the potential energy calculation. However, computational effort is drastically decreased by the applied truncation method; here, the no-truncation would cost N^2 where N is the number of particles.

Several other methods are proposed to minimize both computational effort and inaccuracies, with the most chosen was Ewald Summation [104]. This method would cost at the scales of $\mathcal{O}(N^{3/2})$. However, the scale is still considerably expensive for the large system. Nevertheless, modifying the scheme by employing particle-mesh Ewald summation [105] would reduce the cost in the scale of $\mathcal{O}(N)$. The proposed scheme is effectively beneficial in the system with the size of $10^3 - 10^4$ particles. Moreover, the scales of $\mathcal{O}(N\log N)$ and $\mathcal{O}(N)$ can be attained with the use of particle-particle/particle-mesh(PPPM) [81] and Greengard and Rokhlin [106] method, respectively. These two proposed methods are satisfactorily capable of handling even bigger system in the order of 10^5 particles.

To briefly discuss the method, the next sections are organized as follow. Firstly, the derivation and the implementation of Ewald summation in the Long-Range calculation are presented. Then, the brief discussion of particle-mesh Ewald and the particle-particle/particle-mesh (PPPM) are delivered.

A.2.1 Ewald Summation

Suppose that the system under the consideration consist of N charged particles at vacuum condition. All particles are positioned in the cell with lattice vector $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. The particle's position is described by $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N$ with charges $q_1, q_2, ..q_N$, respectively. To make a more simple description, the system is assumed to be neutral as a whole, in such a way that $\sum_{i=1}^{N} q_i = 0$. Moreover, lets further assume that the system is subjected to a periodic boundary condition in which $\mathbf{r}_i + \mathbf{n}L = \mathbf{r} + n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$ where $n_i, i = 1, 2, 3$ are arbitrary constant. Variable of L is the characteristic length of the cell in which the particle with the same charge at $\mathbf{r}_i$ can also be found at $\mathbf{r}_i + L$. Within these assumption, the potential form of the system excluding the contribution of particle i for the whole space and imposing periodic boundary condition can be written as,

$$\phi_{[i]}(\mathbf{r}) \equiv \phi(\mathbf{r}) - \phi_i(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \sum_{j,\mathbf{n}}{}' \frac{q_j}{|\mathbf{r} - \mathbf{r}_j + \mathbf{n}L|} \ , \tag{A.9}$$

where $'$ sign denotes that $j \neq i$ when $\mathbf{n} = 0$. Then, the total Coulomb electrostatic energy for the whole space can be written as

$$\mathcal{U}_{\text{Coul}} = \frac{1}{2} \sum_{i=1}^{N} q_i \phi_{[i]}(\mathbf{r}) \, . \tag{A.10}$$

However, Eq.[A.10] incorporate poorly (or partially) convergent sum. Concerning this limitation, the method is therefore cannot be used directly in the calculation of electrostatic energy of the system subjected to periodic boundary condition. To handle the problem, Gaussian unit which is compact in notation's form is used. Starting from the following relation,

$$\rho_i(\mathbf{r}) = q_i\delta(\mathbf{r} - \mathbf{r}_i) \tag{A.11}$$

charge density is modified to enhance the convergent. By using this approach, the electrostatic energy form of Eq.[A.10] transformed into the sum of δ - function. In the more general case, the charge density is not always δ function. In the case of the long-range consideration, the charge distribution can also be spread out in space with smoothly varying behavior. For this general problem, the calculation of the electrostatic potential is started from the solving of the following Poisson equation

$$\nabla^2\phi_i(\mathbf{r}) = -\frac{\rho_i(\mathbf{r})}{\varepsilon_0}, \tag{A.12}$$

where the solution can be written as

$$\phi_i(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\mathrm{d}^3\mathbf{r}' \tag{A.13}$$

for the whole space. In the case where the charged particle is distributed with δ-function, the electronic energy can be calculated as follows. By rewriting Eq.[A.10] with the substitution of Eq.[A.13] and considering Eq.[A.11], the energy equation can be written as

$$\mathcal{U}_{\mathrm{Coul}} = \frac{1}{2}\frac{1}{4\pi\varepsilon_0}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}{}' \int\int \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' + \mathbf{n}L|}\,\mathrm{d}^3\mathbf{r}\;\mathrm{d}^3\mathbf{r}' \tag{A.14}$$

and the potential function excluding the contribution of ion i is given as follow

$$\phi_{[i]}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\sum_{\mathbf{n}}\sum_{j=1}^{N}{}' \int \frac{\rho_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' + \mathbf{n}L|}\,\mathrm{d}^3\mathbf{r}' \;. \tag{A.15}$$

The idea of reducing the computational cost is governed as follows. Let's describe the system in a way that each particle is surrounded by the opposite diffused charge. The total charge of this cloudy charge at particle i cancels the particle charge itself, q_i. The diffused charge's distribution is smoothly varied in the space. The calculation of potential now comes from the contribution of the three considerate charges, namely the point, the cloud with the total charge of one enclosure is $-q_i$, and the additional diffusive distributed charge with the total charge of q_i. Due to the main purpose of this calculation is to obtain the potential generated by the collection of point charges of q_i in the whole space, the last mentioned charge's species is needed to compensate the contribution of the previous two. To make a clear description, the illustration of this splitting charged system is depicted in

figure A.1. The three charges species are not necessarily be included in the electrostatic

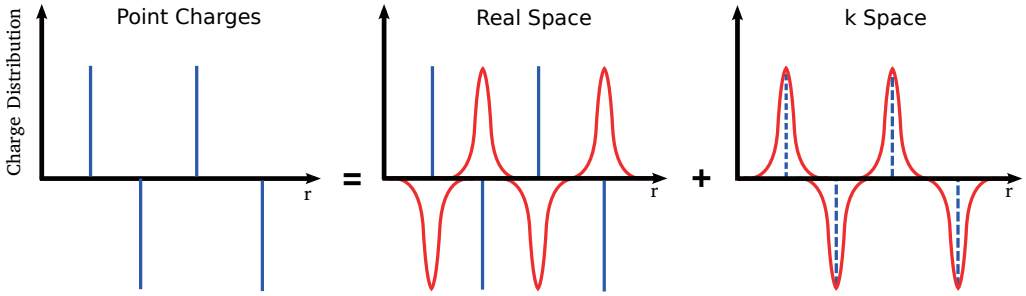


Fig. A.1. Illustration of Ewald Summation

potential calculation at q_i to avoid the self-interaction. However, when the system is subjected to the periodic boundary condition, the compensated charge can exceptionally be included in the calculation of electrostatic potential at q_i. Since the function is periodic and varies smoothly, the q_i involvement can be employed in the calculation. Therefore, the distribution function can be further described by Fourier series to simplify the calculation. As a consequence of this implementation, an energy correction is needed to be performed at the end of the calculation due to the spuriousity of the self-interaction. The following equation

$$\begin{aligned}\rho_i(\mathbf{r}) &= \rho_i^S(\mathbf{r}) + \rho_i^L(\mathbf{r}) \\ \rho_i^S(\mathbf{r}) &= q_i\delta(\mathbf{r}-\mathbf{r}_i) - q_i G_\sigma(\mathbf{r}-\mathbf{r}_i) \\ \rho_i^L(\mathbf{r}) &= q_i G_\sigma(\mathbf{r}-\mathbf{r}_i)\,,\end{aligned} \tag{A.16}$$

of charge density is declared, where

$$G_\sigma(\mathbf{r}) = \frac{1}{(2\pi\sigma^2)^{3/2}}\exp\left[-\frac{|\mathbf{r}|^2}{2\sigma^2}\right]\,, \tag{A.17}$$

and σ is a standard deviation or half of the Gaussian width. By using this form, the following relation

$$\lim_{\sigma\to 0} G_\sigma(\mathbf{r}) = \delta(\mathbf{r})\,, \tag{A.18}$$

is hold. Furthermore, the splitting of a charge generate the following potential function

$$\begin{aligned}\phi_i(\mathbf{r}) &= \phi_i^S(\mathbf{r}) + \phi_i^L(\mathbf{r}) \\ \phi_i^S(\mathbf{r}) &= \frac{q_i}{4\pi\varepsilon_0}\int \frac{\delta(\mathbf{r}-\mathbf{r}') - G_\sigma(\delta\mathbf{r}-\mathbf{r}')}{|(\mathbf{r}-\mathbf{r}')|}\mathrm{d}^3\mathbf{r}' \\ \phi_i^L(\mathbf{r}) &= \frac{q_i}{4\pi\varepsilon_0}\int \frac{G_\sigma(\delta\mathbf{r}-\mathbf{r}')}{|(\mathbf{r}-\mathbf{r}')|}\mathrm{d}^3\mathbf{r}'\,.\end{aligned} \tag{A.19}$$

Rewriting the equation by excluding the contribution of ion i gives

$$\phi_{[i]}(\mathbf{r}) = \phi^S_{[i]}(\mathbf{r}) + \phi^L_{[i]}(\mathbf{r}). \tag{A.20}$$

The splitting will make the Coulomb interaction energy to be written as

$$\mathcal{U}_{\text{Coul}} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi^S_{[i]}(\mathbf{r}_i) + \frac{1}{2}\sum_{i=1}^{N} q_i \phi^L_{[i]}(\mathbf{r}_i)\,. \tag{A.21}$$

In Eq.[A.21], system's energy comes from the contribution of long range interaction. Rewriting this energy equation by including the contribution of self-interactions gives the following equation,

$$\begin{aligned}\mathcal{U}_{\text{Coul}} &= \frac{1}{2}\sum_{i=1}^{N} q_i \phi^S_{[i]}(\mathbf{r}_i) + \frac{1}{2}\sum_{i=1}^{N} q_i \phi^L(\mathbf{r}_i) - \frac{1}{2}\sum_{i=1}^{N} q_i \phi^L_i(\mathbf{r}_i) \\ &\equiv \mathcal{U}^S_{\text{Coul}} + \mathcal{U}^L_{\text{Coul}} - \mathcal{U}^{\text{self}}_{\text{Coul}}\,,\end{aligned} \tag{A.22}$$

where

$$\mathcal{U}^S_{\text{Coul}} \equiv \frac{1}{2}\sum_{i=1}^{N} q_i \phi^S_{[i]}(\mathbf{r}_i) \tag{A.23}$$

$$\mathcal{U}^L_{\text{Coul}} \equiv \frac{1}{2}\sum_{i=1}^{N} q_i \phi^L(\mathbf{r}_i) \tag{A.24}$$

$$\mathcal{U}^{\text{self}}_{\text{Coul}} \equiv \frac{1}{2}\sum_{i=1}^{N} q_i \phi^L_i(\mathbf{r}_i) \tag{A.25}$$

Due to the introduction of smearing function, Coulomb energy function is now can be divided into three main parts, namely long range, short range and a constant self-interaction part as the correction; here, shielding charge clouds and point charge determined the smearing function. Since it can converge rapidly, the short-range part can then be directly calculated in the real space. In contrast, the long-range part should be derived in the reciprocal space.

A.2.2 Potential Field of Gaussian Charge Distribution

To derive the electrostatic potential generated by the Gaussian-distributed charge, the following Poisson equation :

$$\nabla^2 \phi_\sigma(\mathbf{r}) = -q\frac{G_\sigma(\mathbf{r})}{\varepsilon_0}\,, \tag{A.26}$$

should be solved first. In spherical coordinates, Eq.[A.26] can be written as

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\left[r\phi_\sigma(r)\right] = -q\frac{G_\sigma(r)}{\varepsilon_0} \tag{A.27}$$

concerning the magnitude of $\phi_\sigma(r)$ depend only on the distance r. The solution of Eq.[A.27] can be written as

$$\phi_\sigma(r) = \frac{q}{4\pi\varepsilon_0}\mathrm{erf}\left(\frac{r}{\sqrt{2}\sigma}\right), \tag{A.28}$$

where $\mathrm{erf}(z) \equiv \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\mathrm{d}t$. Therefore, Eq.[A.21] can be rewritten as

$$\begin{aligned}\phi_i^S(\mathbf{r}) &= \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\mathrm{erf}\left(\frac{|\mathbf{r}-\mathbf{r}_i|}{\sqrt{2}\sigma}\right)\\ \phi_i^L(\mathbf{r}) &= \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\mathrm{erf}\left(\frac{|\mathbf{r}-\mathbf{r}_i|}{\sqrt{2}\sigma}\right)\end{aligned} \tag{A.29}$$

where $\mathrm{erfc}(z) \equiv 1 - \mathrm{erf(z)}$.

From the equation, since $\lim_{z\to 0}\mathrm{erf(z)} = 1$, the resulted long-range potential is now a non-singular function. In comparison to the Coulomb potential function generated by the point charge in which consist of singular function and long-range function, Eq.[A.29] gives only singular function in short range. Therefore, this function provides potential truncation at long distance interactions. In the following, starting from the contribution of excluded ion i in the short-range calculation that can be written as

$$\phi_{[i]}^S(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\sum_{\mathbf{n}}\sum_{j=1}^{N}{}'\frac{q_j}{|\mathbf{r}-\mathbf{r}_j+\mathbf{n}L|}\,\mathrm{erfc}\left(\frac{|\mathbf{r}-\mathbf{r}_j+\mathbf{n}L|}{\sqrt{2}\sigma}\right), \tag{A.30}$$

and the substitution of Eq.[A.30] to Eq.[A.23] which yield

$$\begin{aligned}\mathcal{U}_{\mathrm{Coul}}^S &\equiv \frac{1}{2}\sum_{i=1}^{N} q_i\phi_{[i]}^S\\ &= \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}{}'\frac{q_iq_j}{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}L|}\,\mathrm{erfc}\left(\frac{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}L|}{\sqrt{2}\sigma}\right)\end{aligned} \tag{A.31}$$

explains the truncation of the short range Coulomb energy calculation by means of $\mathrm{erfc}(z)$. Since the summation is in the real space, Eq.[A.31] can be directly calculated. Meanwhile, the self interaction is now can be derived as follow. Since the following relation

$$\lim_{z\to 0}\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}z\,, \tag{A.32}$$

is hold, rewriting Eq.[A.29] with the use of Eq.[A.32] gives electrostatic potential function

that can be written as

$$\phi_i^L(\mathbf{r}_i) = \frac{q_i}{4\pi\varepsilon_0}\sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\,. \tag{A.33}$$

The self interaction part of coulomb energy function is then given by

$$\mathcal{U}_{\text{Coul}}^{\text{self}} = \frac{1}{4\pi\varepsilon_0}\frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{N} q_i^2\,. \tag{A.34}$$

A.2.3 Long Range Potential in Reciprocal Space

In the previous section, the derivation of potential function and the calculation of electrostatic Coulomb energy has been conducted by introducing the contribution of three charge distributions. These derivation yield the calculation to be split into three main parts, namely short-range, long-range and self-interaction. For the short and self-part, the calculation can be done in the real space. For the long interaction part, the calculation is needed to be done in reciprocal space since the function is no longer singular. In the following discussion, the calculation will be done in the system considered to be subjected into periodic boundary condition. By rewriting Eq.[A.16] for the long-range charge's density of the whole space gives the following relation

$$\rho^L(\mathbf{r}) = \sum_{\mathbf{n}}\sum_{i=1}^{N}\rho_i^L(\mathbf{r}+\mathbf{n}L)\,, \tag{A.35}$$

where can be viewed as a periodic array of ions. Therefore, $\phi^L(\mathbf{r})$ can also be seen to be generated by a periodic array of ions. This form can be brought into the reciprocal space by employing Fourier tranform. Let the following transformation is evaluated in the volume V of supercell

$$\begin{aligned}\hat{\rho}^L(\mathbf{k}) &= \int_V \rho^L(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3\mathbf{r}\\ \hat{\phi}^L(\mathbf{k}) &= \int_V \phi^L(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3\mathbf{r}\end{aligned} \tag{A.36}$$

for $\phi^L\mathbf{r}$ and $\rho^L(\mathbf{r})$, respectively. The inverse transformation can be written as

$$\begin{aligned}\rho^L(\mathbf{r}) &= \frac{1}{V}\sum_{\mathbf{k}}\hat{\rho}^L(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\\ \phi^L(\mathbf{r}) &= \frac{1}{V}\sum_{\mathbf{k}}\hat{\phi}^L(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\,.\end{aligned} \tag{A.37}$$

These two are related to the Poisson equation in which can be written in real space as

$$\nabla^2\phi^L(\mathbf{r}) = -\frac{\rho^L(\mathbf{r})}{\varepsilon_0} \tag{A.38}$$

while in reciprocal space is given as follow

$$k^2 \hat{\phi}^L(\mathbf{k}) = \frac{\hat{\rho}^L(\mathbf{k})}{\varepsilon_0}\,, \tag{A.39}$$

where the solution of reciprocal potential function is just the Fourier transform of charge distribution dividing by k^2. Accordingly, by employing the transformation of long range charge density as can be given in the following

$$\begin{aligned}
\rho^L(\mathbf{r}) &= \sum_{\mathbf{n}} \sum_{j=1}^{N} q_j G_\sigma(\mathbf{r} - \mathbf{r}_j + \mathbf{n}L) \\
\hat{\rho}^L(\mathbf{k}) &= \int_V \sum_{j=1}^{N} q_j G_\sigma(\mathbf{r} - \mathbf{r}_j + \mathbf{n}L) e^{-i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} \\
&= \sum_{j=1}^{N} q_j \int_{\mathbf{R}^3} G_\sigma(\mathbf{r} - \mathbf{r}_j) - i\mathbf{k}\cdot\mathbf{r}\, \mathrm{d}^3\mathbf{r} \\
&= \sum_{j=1}^{N} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} e^{-\sigma^2 \frac{k^2}{2}} \\
&= \sum_{j=1}^{N} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \hat{G}_\sigma(\mathbf{k}) \\
&= \hat{\rho}(\mathbf{k})\, \hat{G}_\sigma(\mathbf{k})
\end{aligned} \tag{A.40}$$

where $k = |\mathbf{k}|$ and $\int_{\mathbf{R}^3}$ is integration over the whole three dimensional space, gives

$$\begin{aligned}
\hat{\rho}_i^L(\mathbf{k}) &= q_i\, \hat{G}_\sigma(\mathbf{k}) \\
\hat{\rho}^L(\mathbf{k}) &= \sum_{i=1}^{N} \hat{\rho}_i^L(\mathbf{k})\,.
\end{aligned} \tag{A.41}$$

This derivation can be dones since $\mathbf{k}$ is a space vector in reciprocal space and $\exp(-i\mathbf{k}\cdot\mathbf{n}) = 1$. The use of $\hat{G}_\sigma(\mathbf{k})$ is to denote the reciprocal diffused function G_σ of charge distribution. This reciprocal space charge density will generate the following potential function

$$\hat{\phi}^L(\mathbf{k}) = \frac{1}{\varepsilon_0} \sum_{j=1}^{N} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \frac{\hat{G}_\sigma(\mathbf{k})}{k^2}\,. \tag{A.42}$$

Subsequently, by employing the inverse Fourier transform to the result gives potential

function in the real space that can be written as

$$\begin{aligned}\phi^L(\mathbf{r}) &= \frac{1}{V}\sum_{\mathbf{k}\neq 0}\hat{\phi}^L(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \frac{1}{V\varepsilon_0}\sum_{\mathbf{k}\neq 0}\sum_{j=1}^{N}\frac{q_j}{k^2}e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_j)}\hat{G}_\sigma(\mathbf{k})\,. \end{aligned} \tag{A.43}$$

When the sum of all charges in the super-cell is neutral, $\sum_{i=1}^{N} q_i = 0$, the contribution to the $\mathbf{k} = 0$ term is zero. By defining the structure factor such as

$$S(\mathbf{k}) \equiv \sum_{i=1}^{N} q_i e^{i\mathbf{k}\cdot\mathbf{r}_i}\,, \tag{A.44}$$

the electrostatic energy function of long range interaction is then can be written as

$$\mathcal{U}^L_{\text{Coul}} = \frac{1}{2V\varepsilon_0}\sum_{\mathbf{k}\neq 0}\frac{\hat{G}_\sigma(\mathbf{k})}{k^2}\left|S(\mathbf{k})\right|^2 \tag{A.45}$$

Finally, the total Coulomb energy of the system can be given as

$$\begin{aligned}\mathcal{U}_{\text{Coul}} &= \mathcal{U}^S_{\text{Coul}} + \mathcal{U}^L_{\text{Coul}} - \mathcal{U}^{\text{self}}_{\text{Coul}} \\ &= \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}{}'\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|}\text{erfc}\left(\frac{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|}{\sqrt{2}\sigma}\right) \\ &+ \frac{1}{2V\varepsilon_0}\sum_{\mathbf{k}\neq 0}\frac{\hat{G}_\sigma(\mathbf{k})}{k^2}\left|S(\mathbf{k})\right|^2 - \frac{1}{4\pi\varepsilon_0}\frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{N}q_i^2\,. \end{aligned} \tag{A.46}$$

A.2.4 Force Derivation

The force derived from the Coulomb interaction energy written in Eq.[A.46] can be given as follow

$$\begin{aligned}\mathbf{f}_j &\equiv -\frac{\partial\mathcal{U}_{\text{Coul}}}{\partial\mathbf{r}_j} \\ &= -\frac{\partial\mathcal{U}^S_{\text{Coul}}}{\partial\mathbf{r}_j} - \frac{\partial\mathcal{U}^L_{\text{Coul}}}{\partial\mathbf{r}_j} \\ &\equiv \mathbf{f}^S_j + \mathbf{f}^L_j\,. \end{aligned} \tag{A.47}$$

There are two main contributions to be considered when calculating the force. First contribution comes from the short range interaction where the potential function of $\phi(r) = \text{erfc}\left[r/(\sqrt{2}\sigma)\right]$ can be computed directly in real space. To handle for the sec-

ond contribution, derivation of the structural factor is first employed as follow

$$\frac{\partial S(\mathbf{k})}{\partial \mathbf{r}_j} = i\,\mathbf{k}\,q_j\,e^{i\mathbf{k}\cdot\mathbf{r}_j}\,, \tag{A.48}$$

to subsequently followed by the following step, where

$$\mathbf{f}_j^L = \frac{1}{2V\varepsilon_0}\sum_{\mathbf{k}\neq 0}\frac{\hat{G}_\sigma(\mathbf{k})}{k^2}\left[S^*(\mathbf{k})\frac{\partial S(\mathbf{k})}{\partial \mathbf{r}_j} + c.c\right], \tag{A.49}$$

comes from the introduction of chain rule. The scheme gives the computational cost scales at $\mathcal{O}\left(N^{3/2}\right)$.

A.2.5 Particle-particle/Particle-Mesh(PPPM)

It is well known that the sum in Coulomb interaction energy function diverges when calculating a large system subjected to periodic boundary condition. To alleviate the problem, in general, an additional form of charge distribution can be employed in a way that convergent potential function can be generated for that such distribution. In more general case, the generated potential function that can be written as

$$\frac{1}{r} = \frac{\mathcal{F}(r)}{r} + \frac{1-\mathcal{F}(r)}{r}\,, \tag{A.50}$$

in which obtained from the separated contributions; here $\mathcal{F}(r)$ is a switching or truncation function.

In the case of Ewald Summation, the splitting produces two kinds of contributions, namely short and long. The generated short-ranged function in potential form comes from the real space erfc and the reciprocal space erf for short and long interaction contribution, respectively. It has been shown that the calculation of Coulomb energy is taking place in two spaces, namely the real and reciprocal space. The first part, namely S part is short-ranged by erfc truncation function when dealing with the long-distance pair interaction. Meanwhile, in the second term namely L, the derived function is non-singular in long-range interaction in which generate short-ranged energy calculation in reciprocal space by $e^{-\sigma^2 k^2/2}$. The less efficient part comes from the Fourier calculation in reciprocal space to cost at $\mathcal{O}(N^2)$ when the cut-off radius is implemented. However, the deficit of computational efficiency comes from the calculation of Poisson equation's solution can be well addressed utilizing the charge distribution in a mesh. Therefore, the chosen function to distribute the charges determined the efficiency scale.

In the case of original Particle-particle/particle-mesh(PPPM) method firstly introduced by Hockney and Eastwood [81], the scheme offered long-range calculation in the scale of $\mathcal{O}(N\log N)$ for the computational cost. This deficiency comes from the implementation of fast Fourier transform (FFT) to solve discretized Poisson equation where interpolated charges on the grid are used. However, this most straightforward implementation give

the energy calculation severe from the accuracy since the splitting procedure written in Eq.[A.50] was not employed yet. Several methods have been conducted to improve the result's accuracy; bringing the essence of Ewald summation into the PPPM scheme was once used to this efficiency recovery. In other words, the calculation is employed by dividing the contribution into two part; here, the particle-mesh scheme is implemented only in the long-range part, and direct calculation on particle-particle interaction is performed in the short range.

In the calculation of the long-range part by using particle mesh scheme, the charge should be placed or interpolated in a grid where the discretized Poisson equation is evaluated. To do this, consider denoting the product of energy as written in Eq.[A.45] into the Fourier space by using Eq.[A.43]. This transformation gives

$$\begin{aligned}\mathcal{U}^{L}_{\mathrm{Coul}} = \frac{1}{2}\sum_{i=1}^{N} q_i \sum_{\mathbf{k}\neq 0} \frac{1}{k^2 V \varepsilon_0} \sum_{j=1}^{N} q_j \\ \times \exp\left[-i\mathbf{k}\cdot\mathbf{r}_j\right] \exp\left[\frac{-\sigma^2 k^2}{2}\right] \exp\left[i\mathbf{k}\cdot\mathbf{r}_i\right] ,\end{aligned} \tag{A.51}$$

where diffused charge distribution follows a Gaussian shape. In more general case where the charge distribution of G_σ can be arbitrary be more diffused and smoothly varied function in real space, Coulomb energy interaction can be written as

$$\begin{aligned}\mathcal{U}^{L}_{\mathrm{Coul}} &= \frac{1}{2}\sum_{i=1}^{N} q_i \sum_{\mathbf{k}\neq 0} \hat{g}(\mathbf{k})\,\hat{\rho}(\mathbf{k})\,\hat{G_\sigma}(\mathbf{k}) \exp\left[i\mathbf{k}\cdot\mathbf{r}_i\right] \\ &= \frac{1}{2}\sum_{i=1}^{N} q_i\,\phi^k(\mathbf{r}_i)\end{aligned} \tag{A.52}$$

where $\hat{g}(\mathbf{k})$ is Green function, $\hat{\rho}(\mathbf{k})$ and $\hat{G_\sigma}(\mathbf{k})$ are reciprocal space form of charge density and smeared function, respectively, in which

$$\phi^k(\mathbf{r}_i) = \frac{1}{V\varepsilon_0}\sum_{\mathbf{k}} \hat{g}(\mathbf{k})\hat{\rho}(\mathbf{k})\,\hat{G_\sigma}(\mathbf{k}\neq 0) \exp\left[i\mathbf{k}\cdot\mathbf{r}_i\right] \tag{A.53}$$

is analogue to the $\frac{1-\mathcal{F}(r)}{r}$ part of Eq.[A.51].

There are several ways proposed by many researchers to give comparable results depending on the system to where the calculation will be carried out. Since the sum of partial charges should be equal the total charge of the system, the function, denoted by $\mathcal{S}(\mathbf{r})$, should be even and normalized. Moreover, the issue that should be addressed when assigning the function is the computational cost gained from many mesh points employed for a single particle charge to distribute. The number of meshes should balance the obtained accuracy that a function can support. Also, last but not least, sudden changes of fractional charges can be generated by the particle's displacement when employing a not suitable function.

In this section, several implementation cases of assigning $\mathcal{S}(\mathbf{r})$ function in a calculation will be briefly explained. To start with, in the case of one dimension problem, a $\mathcal{S}(x)$ weighting function is introduced to assign partial distributed charge on a calculation grid. For a particle at x_r, fraction of charge lies on a grid x_a with density function of $\rho(x) = \sum_i q_i \delta(x - x_i)$ can be given by $\mathcal{S}(x_a - x_r)$, in which

$$\rho_{\mathrm{F}}(x_a) = \frac{1}{h} \int_0^{\mathcal{L}} \mathrm{d}x \mathcal{S}(x_a - x_r)\, \rho(x), \tag{A.54}$$

denote the charge distribution on a grid point of F similar h distance segmented line with total length $\mathcal{L} = h \times \mathrm{F}$.

Correspond to the 3-dimensional problem; charges are assigned to the grid by introducing a function where the charge density at a grid point, $\mathbf{R}$ is [107]

$$\rho(\mathbf{R}) = \frac{q(\mathbf{R})}{\mathcal{L}_x \mathcal{L}_y \mathcal{L}_z}\,, \tag{A.55}$$

The q charge will be placed in the mesh with the cartesian x, y, and z spacing are $\mathcal{L}_x, \mathcal{L}_y$, and $\mathcal{L}_z$, respectively. The author uses 8 surrounding points when placing partially distributed charges where it previously located at the off-grid point. The charges which are distributed to $\mathbf{R}$ mesh-point can be written as

$$q(\mathbf{R}) = \sum_{i=1}^{N} q_i \boldsymbol{\mathcal{S}}(\mathbf{r}_i - \mathbf{R})\,, \tag{A.56}$$

where $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_x, \mathcal{S}_y, \mathcal{S}_z\}$ is a smeared function. This function map the charge previously located at $\mathbf{r}_i$ to a grid of $\mathbf{R}$ that can be written as

$$\mathcal{S}_x(r_{i,x} - R_x) = \begin{cases} 1 - \frac{|r_{i,x}^2 - R_x|}{\mathcal{L}_x}, & \text{if } |r_{i,x}^2 - R_x| < \mathcal{L}_x \\ 0, & \text{if } |r_{i,x}^2 - R_x| \geq \mathcal{L}_x \end{cases}. \tag{A.57}$$

Calculation accuracy can be enhanced by employing more grid points when assigning the charge in this first step. The author further suggested that error would be from the scheme when one is calculating the gradient of $\rho(\mathbf{R})$ and $\phi(\mathbf{R})$ to get electrostatic forces and energies, respectively. Redistribution of charge to a more larger area should be employed to alleviate the flaw. This stage incorporates approximately 500 grid points in a certain enclosed radius of R_c. Accordingly, the new charge distribution at $\mathbf{R}$ in which redistributed from the previous partial charge at $\mathbf{R}_0$ can be written as

$$q\prime(\mathbf{R}) = \begin{cases} q(\mathbf{R_0}) \mathcal{S}_2(|R - R_0|), & \text{if } |R - R_0| < R_c \\ 0, & \text{if } |R - R_0| \geq R_c \end{cases} \tag{A.58}$$

by introducing $\mathcal{S}_2(|R - R_0|)$. This second smeared function would produce a more smooth distribution in a broader areas and beneficial to reduce errors. There are two forms of the

function, namely

$$\mathcal{S}_{\text{linear}}(r) = 1 - \frac{r}{R_c} M^{-1} \,, \quad \text{, and}$$
$$\mathcal{S}_{\text{Gauss}}(r) = \sigma^{-3}\pi^{2/3} \exp\left(-\frac{r^2}{\sigma^2}\right) M^{-1} \tag{A.59}$$

in which comparable results would be produced to address the purpose. The sum of $\mathcal{S}$ within $\mathbf{R}$ is normalized by M^{-1}.

All of the previous weighting function, in the first place, are derived concerning the problem of signing a charge distribution into the calculation grid. However, some authors would argue that the evaluation of Fourier transform can be viewed in a perspective of interpolation problem in the complex exponential part. With regard to the last term of Eq.[A.53], discrete Fourier sum cannot be directly employed in $\mathbf{r}_i$ since it does not coincide with the mesh point. To handle the problem, lets assume that the system is 3 dimensional with edge dimension defined by $\mathcal{L}_\alpha$ are segmented into similar Λ_α grid spaces with length n_α, where $\alpha = \{x, y, z\}$. Lets further assume that an atom at $\mathbf{r}_i$ located between grid points of $\mathrm{u}_{i,\alpha}$ and $\mathrm{u}_{i,\alpha} + 1$ where $\mathrm{u}_{i,\alpha} = \lfloor \frac{\Lambda_\alpha r_{i,\alpha}}{\mathcal{L}_\alpha} \rfloor$. In term of a complex exponential sum in Eq.[A.53], interpolation of order-2p can be written as follow

$$\exp\left[-i\, k_\alpha\, \mathrm{r}_{i,\alpha}\right] \approx \sum_{j=-\infty}^{\infty} \mathcal{S}_{2p}(\mathrm{u}_{i,\alpha} - j) \exp\left\{\frac{-ik\mathcal{L}_\alpha j}{\Lambda_\alpha}\right\} . \tag{A.60}$$

Order 2p interpolation only involving Λ_α terms in the summation of j. This scheme can be further viewed as a Fourier transform of mesh-point charge density. To do that, lets rewrite Eq.[A.43] in such a way correspond to the approximation written above to give

$$\hat{\rho}_k^L \approx \sum_{i=1}^{N} q_i \sum_{j=-\infty}^{\infty} \mathcal{S}_{2p}(\mathrm{u}_{i,\alpha} - j) \exp\left\{\frac{-ik\mathcal{L}_\alpha j}{\Lambda_\alpha}\right\} , \tag{A.61}$$

for the total charge density, and can be further written as

$$\hat{\rho}_k^L \approx \sum_{j} \exp\left\{\frac{-ik\mathcal{L}_\alpha j}{\Lambda_\alpha}\right\} \sum_{i=1}^{N} q_i \mathcal{S}_{2p}(\mathrm{u}_{i,\alpha} - j) \,, \tag{A.62}$$

to slightly shows that $\mathcal{S}_{2p}$ not only give interpolation scheme of complex exponential form, indeed give the way to partially distribute a charge in mesh points.

After the charge distribution scheme was done with many ways as mentioned above, the second stage to do PPPM process is to solve the discretized Poisson equation on a grid in which charges are already partially distributed. When splitting procedure is strictly used, the operation only takes the long-range interaction to calculate. This simplified scheme is available since the short-range uses direct pair calculation in real space. Therefore, the reciprocal space form of charge distribution obtained from Fast Fourier transform is employed when solving the Poisson equation for potential distribution.

The third stage in the calculation scheme of PPPM implemented in molecular dynamics simulation is potential gradient calculation. Force on a grid point should be calculated by taking the gradient of the resulting Coulomb potential. Finally, the obtained forces are then mapped back into the real space to drive the atom displacement.

A.3 Thermostat and Barostat

The thermostat is used in the simulation setup to modulate the temperature. Realistically, applying the thermostat process to get NPT condition in MD simulation is similar to the removal and the additional energy from the boundaries. Instead of getting constant property similarly by devising the value of kinetic energy to be fixed, the purpose of applying this scheme is to ensure the simulated system approximately reach the desired canonical ensemble by allowing the acceptable fluctuations in temperature.

In this study, Nosé-Hoover based algorithm of thermostating and barostating are chosen to generate the canonical ensemble as the scheme ensures the dynamical properties of the system is not having a discontinuity. Extended Lagrangian is applied to derive the fluctuation of the simulated system to the desired temperature condition. Additional artificial coordinates and velocities are added to the Lagrangian. It is proven that this thermostat is stable and efficient since the enhanced optimization is included in each time step.

However, when considering the history of the algorithm, in some instances, a canonical condition in the phase space was not appropriately derived when applying the original Nosé-Hoover algorithm. This flaw happened when external forces are applied to the system. Although for some cases it operates appropriately, several exceptional circumstances contribute to the failure. This exception was identified, for instance, when one applying the scheme to the harmonic oscillator for a given set of initial conditions. The non-ergodic behavior is discovered as the flaw in producing canonical ensemble and the dependency on the initial configurations. It was later explained that the energy is not the only conserved quantity as it was in the system where no external forces applied, and the center of mass remains fixed. Center-of-mass can be adjusted to zero as a solution for keeping the scheme working correctly, but this is not a solution for the more general cases.

Martyna et al. [79] suggested another conservation law in the scheme to alleviate the limitations. They coupled another thermostat to the Nosé thermostat and further connecting to a chain of thermostats if it is needed. These chains are accounted as the additional conservation law. It was proved by applying the scheme to the original Nosé-Hoover algorithm; the modified counterpart produced proper canonical ensemble.

Experimental setup mostly is done under the constant pressure and temperature condition. These setups are prepared when one needs to observe the samples in such a way that constant pressure is required for the fascinating behavior to be found. To elaborate on these conditions in the simulation process, one needs to preserve additional quantity, which is pressure, by adjusting the volume from NVT condition. In doing this, the volume

is considered as a dynamical variable changing with time during simulations. Since the Martyna et al. [79] scheme produce the proper NVT conditions, additional treatments are needed to ensure further the algorithm delivers the correct NPT ensemble. Andersen's scheme introduced previously offered the algorithm that can be applied to the Martyna's generalization of Nosé-Hoover original thermostat scheme. It is proved that the modified method produced the correct ensemble.

When one considers studying solid and amorphous polymers by using molecular dynamics, it is often essential to include the presence of external stress. Finding mechanical strength by preserving the stress-strain behavior and elastic constant is beneficial to determine the stability criteria. To do this, Parrinello-Rahman has established the molecular dynamics of the NtH ensemble, allowing the system with N particles to change the shape of the cell during simulation. The t stand for thermodynamics tension while H for enthalpy. However, straightforward approach of elaborating this molecular dynamics scheme into the process dealt with time-consuming problems as consequences even for the zero-stress case. It takes millions of steps to get a satisfactory convergence result.

Shinoda et al. [78] introduced the NtT scheme of MD simulation by applying the efficient sampling techniques. The non-hamiltonian equation of motion is integrated by using the rRESPA algorithm to enables the various thermostating process such as original Nosé-Hoover scheme, Nosé-Hoover chain and Massive Nosé-Hoover chain (MNHC) to be implemented. This scheme was proved faster and properly produce the desired ensembles compared to the previous, thus being considered as the solution for the problems regarding the high cost computational time grows in the simulating systems where stress is applied while other solution failed in handling error inevitably, not working for the generally established force field and statistical convergence problems. Additionally, the time reversibility on this algorithm allowing one to implement in hybrid Monte Carlo simulation. In this study, the discussion of thermostat and barostat is started from the derivation of the equation of motion of the simulated system. The formulation of original Nosé-Hoover equation motion is also explained. Subsequently, chain and massive Nosé-Hoover thermostats will be derived and explained from the original thermostat scheme. Finally, the barostat is explained followed by the introduction of the rRESPA integration.

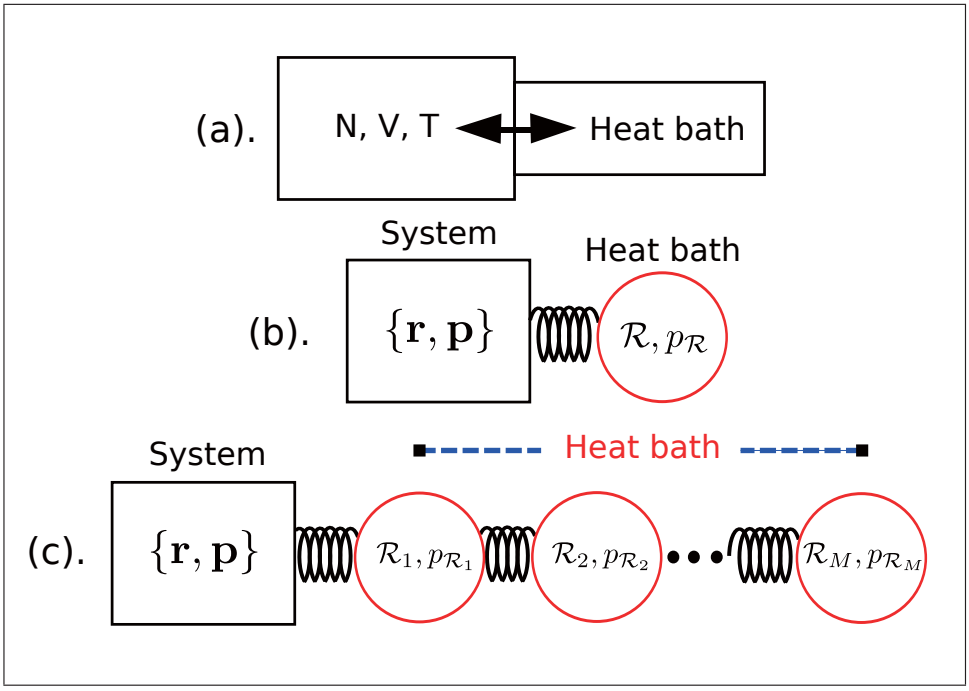


Fig. A.2. Illustration of Thermostating in Molecular dynamics for (a). General description, (b). Nosé-Hoover method, and (c). Nosé-Hoover chain method

A.3.1 Nosé-Hoover Algorithm

Suppose that in the simulated system exist N particles which are describes by position vector $\mathbf{r}_i$ and momentum $\mathbf{p}_i$. The Nosè-Hoover equation of motion is written as follow:

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_{\mathcal{R}}}{\mathcal{Q}}\mathbf{p}_i \\
\dot{\mathcal{R}} &= \frac{p_{\mathcal{R}}}{\mathcal{Q}} \\
\dot{p}_{\mathcal{R}} &= \sum_{i=1}^{N} \frac{p_i^2}{m_i} - g k_B T \,,
\end{aligned} \tag{A.63}$$

where g is a parameter to properly determined in order to produce canonical distribution.

Next, conservation laws and the non-driven variables should be derived by analyzing the system dynamics. The Nosé hamiltonian assuming the energy conservation can be written as follow

$$\begin{aligned}
\mathcal{H}_{\mathrm{Nos'e}} &= \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \phi(\mathbf{r}^N) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + g k_B T_{\mathcal{R}} \\
&= \mathcal{H}(\mathbf{r}, \mathbf{p}) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + g k_B T \mathcal{R} = C_1 \,,
\end{aligned} \tag{A.64}$$

where $\mathcal{H}(\mathbf{r},\mathbf{p})$ is a physical Hamiltonian of the system. To derive the system's compressibility, variable of $\boldsymbol{\eta} = \left(\mathbf{r}^N, \mathbf{p}^N, \mathcal{R}, p_{\mathcal{R}}\right)$ is introduce to be written in the form as follow

$$\begin{aligned}\kappa(\boldsymbol{\eta}) &= \boldsymbol{\nabla}_{\boldsymbol{\eta}} \cdot \dot{\boldsymbol{\eta}} \\ &= \sum_{i=1}^{N} \boldsymbol{\nabla}_{\mathbf{r}_i} \cdot \dot{\mathbf{r}}_i + \sum_{i=1}^{N} \boldsymbol{\nabla}_{\mathbf{p}_i} \cdot \dot{\mathbf{p}}_i + \boldsymbol{\nabla}_{\mathcal{R}} \cdot \dot{\mathcal{R}} + \boldsymbol{\nabla}_{p_{\mathcal{R}}} \cdot \dot{p_{\mathcal{R}}} \\ &= \sum_{i=1}^{N} \boldsymbol{\nabla}_{\mathbf{p}_i} \cdot \dot{\mathbf{p}}_i \\ &= -\frac{dN p_{\mathcal{R}}}{\mathcal{Q}} = -dN\dot{\mathcal{R}}\,. \end{aligned} \tag{A.65}$$

The system's Jacobian resulted from the Eq.[A.65] can be written as

$$J = \exp\left[-\int \kappa(\boldsymbol{\eta})\mathrm{d}t\right] = \exp(dN\mathcal{R})\,. \tag{A.66}$$

Let us previously reconsider the purpose of applying the algorithm which will resulted the proper NVT ensemble. Rewriting the partition function of the non-Hamiltonian system NVT ensemble for n_c conservation laws and $\Lambda_k(\boldsymbol{\eta}') = C_k$ for $k = 1, 2, ..., n_c$ yields

$$\Omega_T(C_1, C_2, C_3, ..., C_n) = \int \mathrm{d}\boldsymbol{\eta}' J(\boldsymbol{\eta}') \prod_{k=1}^{n_c} \delta(\Lambda_k(\boldsymbol{\eta}') - C_k)\,. \tag{A.67}$$

By substituting the metric in Eq.[A.65] to the partition function of Eq.[A.67] gives

$$\begin{aligned}\Omega_T(N, V, C_1) = \int \mathrm{d}^N\mathbf{p} \int \mathrm{d}^N\mathbf{r} \int \mathrm{d}p_{\mathcal{R}} \int \mathrm{d}\mathcal{R} \times \exp(dN\mathcal{R}) \\ \delta\left(\mathcal{H}(\mathbf{r},\mathbf{p}) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\mathcal{R} - C_1\right)\,. \end{aligned} \tag{A.68}$$

The integration over the extended variables of$\mathcal{R}$ and $p_{\mathcal{R}}$, can be done analytically; here, the variables are used to describe thermostat's position and momentum, respectively. To employ the process, rewriting $\mathcal{R}$ variable to handle the δ function as given below

$$\mathcal{R} = \frac{1}{gk_BT}\left[C_1 - \mathcal{H}(\mathbf{r},\mathbf{p}) - \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}}\right]\,, \tag{A.69}$$

and substituting the result to Eq.[A.68] gives

$$\begin{aligned}\Omega_T(N,V,C_1) &= \frac{1}{gk_BT}\int \mathrm{d}^N\mathbf{p}\int \mathrm{d}^N\mathbf{r}\int \mathrm{d}p_{\mathcal{R}}\int \mathrm{d}\mathcal{R} \\ &\times \exp\left[\frac{dN}{gk_BT}\left(C_1 - \mathcal{H}(\mathbf{r},\mathbf{p}) - \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}}\right)\right] \\ &= \frac{\exp\left(dNC_1/gk_BT\right)}{gk_BT}\int \mathrm{d}p_{\mathcal{R}}\exp\left(-\frac{dNp_{\mathcal{R}}^2}{2Qg}\right) \\ &\times \int \mathrm{d}^N\mathbf{p}\int \mathrm{d}^N\mathbf{r}\,\exp\left(-\beta\frac{\mathrm{d}N\mathcal{H}(\mathbf{r},\mathbf{p})}{g}\right). \end{aligned} \tag{A.70}$$

When we make g equal to dN, then the following relation,

$$\Omega_T(N,V,C_1) \propto Q(N,V,T)\,, \tag{A.71}$$

holds. Integration of other extended variables $p_{\mathcal{R}}$ can be physically negligible as a result only a constant. According to the system under evaluation, the derivation of this scheme resulting from only one conservation law as shown in Eq.[A.64]. In general, as previously explained, this scheme will fail to produce the correct canonical ensemble in certain cases such as a harmonic oscillator. To discuss the more general case in which needed additional conservation law, suppose that we simulate a system where the total momentum is conserved as a consequence of no external forces being applied. As can be shown starting from the following equation, this condition will affect phase-space distribution. Conservation of total momentum in the view of Nosé-Hoover dynamics can be given by

$$\begin{aligned}\frac{\mathrm{d}\mathbf{P}\mathrm{e}^{\mathcal{R}}}{\mathrm{d}t} &= \mathrm{e}^{\mathcal{R}}\left(\dot{\mathbf{P}} + \mathbf{P}\dot{\mathcal{R}}\right) = \mathrm{e}^{\mathcal{R}}\left(\dot{\mathbf{P}} + \mathbf{P}\frac{p_{\mathcal{R}}}{\mathcal{Q}}\right) \\ &= \mathrm{e}^{\mathcal{R}}\left(\sum_{i=1}^{N}\left(\mathbf{F}_i - \frac{p_{\mathcal{R}}}{\mathcal{Q}}\mathbf{p}_i\right) + \mathbf{P}\frac{p_{\mathcal{R}}}{\mathcal{Q}}\right) \\ &= 0, \end{aligned} \tag{A.72}$$

to yields :

$$\mathbf{P}\mathrm{e}^{\mathcal{R}} = \mathbf{K}\,, \tag{A.73}$$

where $\mathbf{P} = \sum_{i=1}^{N}\mathbf{p}_i$ is the center-of-mass momentum while $\mathbf{K}$ is an arbitrary vector. In Eq.[A.64] we have the magnitude of center-of-mass momentum $\mathbf{P}$ coupled to other variables that can not be eliminated for further analysis; here, the momentum is a driven variable. However, by considering the following equation

$$\frac{P_x}{K_x} = \frac{P_y}{K_y} = \frac{P_z}{K_z} = e^{\mathcal{R}}, \tag{A.74}$$

we can see that the component of momentum $\mathbf{P}$ is linearly dependent. Based on this, one of the momentum $\mathbf{P}$ component can be taken as $P = (\sum_\alpha P_\alpha^2)^{1/2}$; this P is the

independent variable. Another driven variables which do not influence the other should be eliminated for further analysis; here, the driven variable is the center-of-mass position. Next, the rewriting of Eq.[A.63] by introducing the position's vector of $\mathbf{r}'$ and momentum of $\mathbf{p}'$ relative to the system's center-of-mass gives

$$\begin{aligned}
\dot{\mathbf{r}}'_i &= \mathbf{p}'_i/m'_i \\
\dot{\mathbf{p}}'_i &= \mathbf{F}'_i - \frac{p_{\mathcal{R}}}{\mathcal{Q}}\mathbf{p}'_i \\
\dot{P}_i &= -\frac{p_{\mathcal{R}}}{\mathcal{Q}}P_i \\
\dot{\mathcal{R}} &= \frac{p_{\mathcal{R}}}{\mathcal{Q}} \\
\dot{p}_{\mathcal{R}} &= \sum_{i=1}^{N-1}\frac{p'^2_i}{m'_i} + \frac{P^2}{M} - gk_BT.
\end{aligned} \tag{A.75}$$

The Hamiltonian can be written as

$$\begin{aligned}
\mathcal{H}_{\mathrm{Nos'e}} &= \sum_{i=1}^{N}\frac{\mathbf{p}'^2_i}{2m'_i} + \frac{P^2}{2M} + \phi(\mathbf{r}'^N) + \frac{p^2_{\mathcal{R}}}{2\mathcal{Q}} + gk_BT\mathcal{R} \\
&= \mathcal{H}(\mathbf{r}', \mathbf{p}', P) + \frac{p^2_{\mathcal{R}}}{2\mathcal{Q}} + gk_BT\mathcal{R} \\
&= \mathcal{H}(\mathbf{r}, \mathbf{p}) + \frac{p^2_{\mathcal{R}}}{2\mathcal{Q}} + gk_BT\mathcal{R} = C_1,
\end{aligned} \tag{A.76}$$

as the first conservation law, and

$$P\exp(\mathcal{R}) = C_2\,, \tag{A.77}$$

is the second. The compressibility equation that can be written as

$$\begin{aligned}
\kappa &= \sum_{i=1}^{N-1}\boldsymbol{\nabla}_{\mathbf{r}'_i}\cdot\dot{\mathbf{r}}'_i + \sum_{i=1}^{N-1}\boldsymbol{\nabla}_{\mathbf{p}'_i}\cdot\dot{\mathbf{p}}'_i + \boldsymbol{\nabla}_P\cdot\dot{P} + \boldsymbol{\nabla}_{\mathcal{R}}\cdot\dot{\mathcal{R}} + \boldsymbol{\nabla}_{p_{\mathcal{R}}}\cdot\dot{p_{\mathcal{R}}} \\
&= -[d(N-1)+1]\dot{\mathcal{R}}\,,
\end{aligned} \tag{A.78}$$

is used to obtain the Jacobian equation as written in the following

$$J = \exp\left\{[d(N-1)+1]\mathcal{R}\right\}. \tag{A.79}$$

Finally, the resulting partition function can be written as

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}} \int \mathrm{d}\mathcal{R} \\ &\times \exp\left\{[d(N-1)+1]\mathcal{R}\right\} \\ &\delta\left(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\mathcal{R} - C_1\right) \\ &\delta\left(\mathrm{e}^{\mathcal{R}}P - C_2\right) . \end{aligned} \tag{A.80}$$

The two conservation laws exist in the partition function as two δ functions. Let us consider the last δ function. As in the following, evaluating the integral of delta imposes that

$$\begin{aligned} e^{\mathcal{R}}P &= C_2 \\ \mathcal{R} &= \ln(C_2/P)\,. \end{aligned} \tag{A.81}$$

Substituting to Eq.[A.80] gives:

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = &\frac{1}{C_2}\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}} \\ &\times \left(\frac{C_2}{P}\right)^{d(N-1)+1} \\ &\delta\left(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\ln(C_2/P) - C_1\right) , \end{aligned} \tag{A.82}$$

for the last δ. Similarly, evaluating the other δ derived from the first conservation law

$$\begin{aligned} \mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\ln(C_2/P) &= C_1 \\ \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} &= C_1 - \mathcal{H}(\mathbf{r}',\mathbf{p}',P) - gk_BT\ln(C_2/P) \\ p_{\mathcal{R}} = \left\{2\mathcal{Q}\left[C_1 - \mathcal{H}(\mathbf{r}',\mathbf{p}',P) - gk_BT\ln(C_2/P) - C_1\right]\right\}^{-\frac{1}{2}} , \end{aligned} \tag{A.83}$$

and upon integrating yields

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = &\frac{\sqrt{2\mathcal{Q}}}{C_2}\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \\ &\times \int \mathrm{d}P\Big(\frac{C_2}{P}\Big)^{d(N-1)+1} \\ &\Big[C_1 - \mathcal{H}(\mathbf{r}',\mathbf{p}',P) - gk_BT\ln(C_2/P)\Big]^{-\frac{1}{2}} . \end{aligned} \tag{A.84}$$

However, Eq.[A.84] is not a partition function of NVT distribution. By considering the second conservation law and set the constant to be $C_2 = 0$, Eq.[A.84] would produce the

following partition function

$$\begin{aligned}\Omega_T(N,V,C_1,0) = &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}} \int \mathrm{d}\mathcal{R} \\ &\times \exp\left\{[d(N-1)+1]\mathcal{R}\right\} \\ &\delta\left(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\mathcal{R} - C_1\right) \\ &\delta\Big(e^{\mathcal{R}}P - 0\Big)\,. \end{aligned} \tag{A.85}$$

This clearly imposes that $P = 0$ when evaluating integration over P as given in the following

$$\begin{aligned}\Omega_T(N,V,C_1,0) = &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}} \int \mathrm{d}\mathcal{R} \\ &\times \exp\left\{[d(N-1)+1]\mathcal{R}\right\} \exp(-\mathcal{R}) \\ &\delta\left(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} + gk_BT\mathcal{R} - C_1\right), \end{aligned} \tag{A.86}$$

while the other δ function imposes

$$\begin{aligned} gk_BT\mathcal{R} &= C_1 - \mathcal{H}(\mathbf{r}',\mathbf{p}',P) - \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} \\ \mathcal{R} &= \frac{1}{gk_BT}\left\{C_1 - \mathcal{H}(\mathbf{r}',\mathbf{p}',P) - \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}}\right\} \\ \mathcal{R} &= -\frac{\beta}{g}\left\{\mathcal{H}(\mathbf{r}',\mathbf{p}') + \frac{p_{\mathcal{R}}^2}{2\mathcal{Q}} - C_1\right\}, \end{aligned} \tag{A.87}$$

which is fixes $\mathcal{R}$. The partition function can be written as

$$\begin{aligned}\Omega_T(N,V,C_1,0) = &\frac{\exp[\beta d(N-1)C_1/g]}{gk_BT} \\ &\times \int \mathrm{d}p_{\mathcal{R}} \exp\left[-\beta d(N-1)p_{\mathcal{R}}^2/(2Qg)\right] \\ &\times \int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \\ &\exp\left[-\beta\frac{d(N-1)}{g}H(\mathbf{r}',\mathbf{p}')\right] \\ \propto &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \\ &\exp\left[-\beta\frac{d(N-1)}{g}H(\mathbf{r}',\mathbf{p}')\right], \end{aligned} \tag{A.88}$$

which will resulted a canonical distribution if one substitute $g = d(N-1)$. It is shown that in original algorithm of Nosé, implementation by fixing system's center-of-mass will resulted $P = 0$ and producing the correct canonical distribution.

A.3.2 Nosé-Hoover Chains

From the previous section, it is shown that when dealing with a system having more than one conservation law, original Nosé algorithm gives partition function which is not canonical distribution. To handle the problem, Martyna et al. [79] introduced other extended variables in the equation of motion. The idea of this more variable's introduction is to coupled single or multiple thermostats to the dynamics, with only having two additional degrees of freedom. Rewriting the equation of motion of N particles attached to M chain of thermostats as given below :

$$\begin{aligned}\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_{\mathcal{R}_1}}{\mathcal{Q}_1}\mathbf{p}_i \\ \dot{\mathcal{R}}_k &= \frac{p_{\mathcal{R}_k}}{\mathcal{Q}_k} \quad k = 1,\cdots,M \\ \dot{\mathcal{R}}_1 &= \left(\sum_{i=1}^{N}\frac{p_i^2}{m_i} - gk_BT\right) - \frac{p_{\mathcal{R}_2}}{\mathcal{Q}_2}p_{\mathcal{R}_1} \\ \dot{p}_{\mathcal{R}_k} &= \left[\frac{p_{\mathcal{R}_{k-1}}^2}{\mathcal{Q}_{k-1}} - k_BT\right] - \frac{p_{\mathcal{R}_{k+1}}}{\mathcal{Q}_{k+1}}p_{\mathcal{R}_k} \\ \dot{p}_{\mathcal{R}_M} &= \left[\frac{p_{\mathcal{R}_{M-1}}^2}{\mathcal{Q}_{M-1}} - k_BT\right], \end{aligned} \tag{A.89}$$

gives

$$H_{\mathrm{NHC}} = \mathcal{H}(\mathbf{r},\mathbf{p}) + \sum_{k=1}^{M} \frac{p_{\mathcal{R}_k}^2}{2\mathcal{Q}_k} + gk_BT\mathcal{R}_1 + \sum_{k=2}^{M} k_BT\mathcal{R}_k\,, \tag{A.90}$$

which is the conserved Hamiltonian. Two additional degrees of freedom come from $\mathcal{R}_1$ and $\mathcal{R}_c = \sum_{l=2}^{M}\mathcal{R}_k$ as the two independently coupled to the system. Suppose that we attach two chains to the system where no external forces applied. As shown previously in Eq.[A.76] and A.77, the same systems will produce two conserve quantities. Firstly, Hamiltonian of the new system in term of the same center-of-mass position and momentum $\{\mathbf{r}',\mathbf{p}'\}$ and total momentum P coordinate system can be written as follow:

$$H_{\mathrm{Nos'e}} = \sum_{i=1}^{N-1} \frac{p_i'^2}{2m_i'} + \frac{P^2}{2M} + \phi\big(\mathbf{r}'^N\big) + \frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1} + \frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c}\,. + gk_BT\mathcal{R}_1 + gk_BT\mathcal{R}_c. \tag{A.91}$$

Simplifying by declaring

$$\mathcal{H}(\mathbf{r}',\mathbf{p}',P) = \sum_{i=1}^{N-1} \frac{p_i'^2}{2m_i'} + \frac{P^2}{2M} + \phi\big(\mathbf{r}'^N\big) = \mathcal{H}(\mathbf{r},\mathbf{p}) \tag{A.92}$$

and further implemented to Eq.[A.91] which can be written as

$$\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1} + \frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c} + gk_BT\mathcal{R}_1 + gk_BT\mathcal{R}_c = C_1\,, \tag{A.93}$$

resulting the first conserved quantity in this case. The second conserved quantity is

$$P\exp(\mathcal{R}_1) = C_2\,. \tag{A.94}$$

Similarly, before continuing to find the metric equation, the compressibility that can be written as

$$\begin{aligned}\kappa &= \sum_{i=1}^{N-1} \boldsymbol{\nabla}_{\mathbf{r}_i'}\cdot\dot{\mathbf{r}}_i' + \sum_{i=1}^{N-1} \boldsymbol{\nabla}_{\mathbf{p}_i'}\cdot\dot{\mathbf{p}}_i' + \boldsymbol{\nabla}_{\mathbf{P}}\cdot\dot{\mathbf{P}} + \boldsymbol{\nabla}_{\mathcal{R}}\cdot\dot{\mathcal{R}} + \boldsymbol{\nabla}_{p_{\mathcal{R}}}\cdot\dot{p_{\mathcal{R}}} \\ &= 0 - d(N-1)\frac{p_{\mathcal{R}_1}}{\mathcal{Q}_1} - \frac{p_{\mathcal{R}_1}}{\mathcal{Q}_1} + 0 - \frac{p_{\mathcal{R}_c}}{\mathcal{Q}_c} \\ &= -(dN-d+1)\dot{\mathcal{R}} - \dot{\mathcal{R}}_c\,,\end{aligned} \tag{A.95}$$

gives the Jacobian as follow

$$J = \exp\left\{[dN-d+1]\mathcal{R}_1 + \mathcal{R}_c\right\}. \tag{A.96}$$

Finally, the partition function for the Nosé-Hoover chain is written as

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}_1} \int \mathrm{d}\mathcal{R}_1 \\ &\int \mathrm{d}p_{\mathcal{R}_c} \int \mathrm{d}\mathcal{R}_c \times \exp\left\{[dN-d+1]\mathcal{R}_1 + \mathcal{R}_c\right\} \\ &\delta\Big(e^{\mathcal{R}_1}P - C_2\Big) \\ &\delta\Bigg(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1} + \frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c} \\ &\quad + gk_BT\mathcal{R}_1 + k_BT\mathcal{R}_c - C_1\Bigg). \end{aligned} \tag{A.97}$$

Delta function comes from the second conserve quantity gives

$$\mathcal{R}_1 = \ln\Big(\frac{C_2}{P}\Big). \tag{A.98}$$

Evaluating the integral by the substitution of Eq.[A.98] to the partition function gives

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = &\int \mathrm{d}^{N-1}\mathbf{p}' \int \mathrm{d}^{N-1}\mathbf{r}' \int \mathrm{d}P \int \mathrm{d}p_{\mathcal{R}_1} \int \mathrm{d}\mathcal{R}_1 \\ &\int \mathrm{d}p_{\mathcal{R}_c} \int \mathrm{d}\mathcal{R}_c \times \Big(\frac{C_2}{P}\Big)^{dN-d+1} \exp(\mathcal{R}_c) \\ &\delta\Bigg(\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1} + \frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c} + gk_BT\ln(C_2/P) \\ &\quad + k_BT\mathcal{R}_c - C_1\Bigg). \end{aligned} \tag{A.99}$$

Evaluating the first conserved quantity appeared in δ function after integrating the first gives

$$\mathcal{R}_c = -\beta\left[\mathcal{H}(\mathbf{r}',\mathbf{p}',P) + \frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1} + \frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c} + gk_BT\ln(C_2/P) - C_1\right]. \tag{A.100}$$

After this step, the partition function can be written as

$$\begin{aligned}\Omega_T(N,V,C_1,C_2) = \frac{\exp(C_1/k_BT)}{k_BTC_2}\int \mathrm{d}^{N-1}\mathbf{p}'\int \mathrm{d}^{N-1}\mathbf{r}'\int \mathrm{d}P\int \mathrm{d}p_{\mathcal{R}_1}\\ \int \mathrm{d}\mathcal{R}_1\int \mathrm{d}p_{\mathcal{R}_c}\int \mathrm{d}\mathcal{R}_c\Big(\frac{C_2}{P}\Big)^{dN-d+1}\\ \exp\Big(-\beta\mathcal{H}(\mathbf{r}',\mathbf{p}',P)\Big)\exp\left\{-\beta\left[\frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1}+\frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c}\right]\right\}\\ \times\exp\Big[-\beta g k_BT\ln(C_2/P)\Big]\\ = \int \mathrm{d}p_{\mathcal{R}_1}\int \mathrm{d}p_{\mathcal{R}_c}\exp\left\{-\beta\left[\frac{p_{\mathcal{R}_1}^2}{2\mathcal{Q}_1}+\frac{p_{\mathcal{R}_c}^2}{2\mathcal{Q}_c}\right]\right\}\\ \int \mathrm{d}^{N-1}\mathbf{p}'\int \mathrm{d}P\times\int \mathrm{d}^{N-1}\mathbf{r}'\exp\Big[-\beta\mathcal{H}(\mathbf{r}',\mathbf{p}',P)\Big]\\ \Big(\frac{P}{C_2}\Big)^{g-d(N-1)-1},\end{aligned} \tag{A.101}$$

in which recovered the canonical distribution if one choosing $g = dN$.

A.3.3 Massive Nosé-Hoover chain thermostat

The scheme offers actively controlled temperature through the attachment of thermostat chains to 3N system's degrees of freedom. As a consequence, a variety of vibration modes can be evident in the system with many frequencies. However, since the system massively coupled to the thermostats, choosing the appropriate thermostat masses can control that spurious behavior utilizing the alleviation of the generated frequency characteristics mimicking the Fourier series.

To briefly discuss the scheme, lets assume that the system is coupled to M thermostat chains with length L on each degree of freedom. The equation of motion can be written as

$$\begin{aligned}\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i}\\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \mathbf{p}_i\sum_{l=1}^{L}\frac{\mathbf{p}_{i,1,l}}{\mathcal{Q}_{i,1,l}}\\ \dot{\boldsymbol{\mathcal{R}}}_{i,k,l} &= \frac{\mathbf{p}_{i,k,l}}{\mathcal{Q}_{i,k,l}}\\ \dot{\mathbf{p}}_{i,1,l} &= \left(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}-k_BT\right)-\mathbf{p}_{i,1,l}\frac{\mathbf{p}_{i,2,l}}{\mathcal{Q}_{i,2,l}}\\ \dot{\mathbf{p}}_{i,k,l} &= \left(\frac{\mathbf{p}_{i,k-1,l}^2}{\mathcal{Q}_{i,k-1,l}}-k_BT\right)-\mathbf{p}_{i,k,l}\frac{\mathbf{p}_{i,k+1,l}}{\mathcal{Q}_{i,k+1,l}}\qquad(2\le k\le M-1)\\ \dot{\mathbf{p}}_{i,M,l} &= \frac{\mathbf{p}_{i,M-1,l}^2}{\mathcal{Q}_{i,M-1,l}}-k_BT\,.\end{aligned} \tag{A.102}$$

Subscripts i, k, and l denote the attached variable correspond to the $l-$th thermostat in chain k of atom i. The conserved total energy can be written as

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \phi(\mathbf{r}^N) + \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\left(\frac{\mathbf{p}_{i,k,l}^2}{2\mathcal{Q}_{i,k,l}} + k_B T \boldsymbol{\mathcal{R}}_{i,k,l}\right), \tag{A.103}$$

since the derivation shown by the following equation

$$\begin{aligned}
\dot{\mathcal{H}} &= \sum_{i=1}^{N}\left(\frac{\partial \mathcal{H}}{\partial \mathbf{p}_i}\dot{\mathbf{p}}_i + \frac{\partial \mathcal{H}}{\partial \mathbf{r}_i}\dot{\mathbf{r}}_i\right) \\
&+ \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\frac{\partial \mathcal{H}}{\partial \mathbf{p}_{i,k,l}}\dot{\mathbf{p}}_{i,k,l} + \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\frac{\partial \mathcal{H}}{\partial \mathcal{R}_{i,k,l}}\dot{\mathcal{R}}_{i,k,l} \\
&= \sum_{i=1}^{N}\left\{\frac{\mathbf{p}_i}{m_i}\left(\mathbf{F}_i - \mathbf{p}_i\sum_{l=1}^{L}\frac{\mathbf{p}_{i,1,l}}{\mathcal{Q}_{i,1,l}}\right) + \mathbf{F}_i\frac{\mathbf{p}_i}{m_i}\right\} \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L}\frac{\mathbf{p}_{i,1,l}}{\mathcal{Q}_{i,1,l}}\left(\frac{\mathbf{p}_i^2}{m_i} - k_B T - \mathbf{p}_{i,1,l}\frac{\mathbf{p}_{i,2,l}}{\mathcal{Q}_{i,2,l}}\right) \\
&+ \sum_{i=1}^{N}\sum_{k=2}^{M-1}\sum_{l=1}^{L}\left\{\frac{\mathbf{p}_{i,k,l}}{\mathcal{Q}_{i,k,l}}\left(\frac{\mathbf{p}_{i,k-1,l}^2}{\mathcal{Q}_{i,k-1,l}} - k_B T - \mathbf{p}_{i,k,l}\frac{\mathbf{p}_{i,k+1,l}}{\mathcal{Q}_{i,k+1,l}}\right)\right\} \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L}\frac{\mathbf{p}_{i,M,l}}{\mathcal{Q}_{i,M,l}}\left(\frac{\mathbf{p}_{i,M-1,l}^2}{\mathcal{Q}_{i,M-1,l}} - k_B T\right) \\
&+ \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}k_B T\frac{\mathbf{p}_{i,k,l}}{\mathcal{Q}_{i,k,l}} = 0,
\end{aligned} \tag{A.104}$$

holds. By assuming that the ergodicity holds and denoting $\mathcal{H} = C$, the approximation of partition function can then be written as

$$\Omega_T \propto \int \mathrm{d}\mathbf{r}^N \int \mathrm{d}\mathbf{p}^N \int \mathbf{d}\boldsymbol{\mathcal{R}}^{NLM} \int \mathrm{d}\mathbf{p}^{NLM} J\delta(\mathcal{H} - C)\,. \tag{A.105}$$

In the same manner of the previous derivation, the Jacobian of such N particle system satisfies the following relation

$$\begin{aligned}
\dot{J} &= -J\sum_{i=1}^{N}\left\{\left(\frac{\partial \dot{\mathbf{r}}_i}{\partial \mathbf{r}_i} + \frac{\partial \dot{\mathbf{p}}_i}{\partial \mathbf{p}_i}\right) + \sum_{k=1}^{M}\sum_{l=1}^{L}\left(\frac{\partial \dot{\boldsymbol{\mathcal{R}}}_{i,k,l}}{\partial \boldsymbol{\mathcal{R}}_{i,k,l}} + \frac{\partial \dot{\mathbf{p}}_{i,k,l}}{\partial \mathbf{p}_{i,k,l}}\right)\right\} \\
&= J\left(\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\frac{\mathbf{p}_{i,k,l}}{\mathcal{Q}_{i,k,l}}\right) = \left(\sum_{i=1}^{N}\sum_{l=1}^{L}\sum_{m=1}^{M}\dot{\boldsymbol{\mathcal{R}}}_{i,k,l}\right)J,
\end{aligned} \tag{A.106}$$

thus delivers,

$$J = \exp\left(\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\boldsymbol{\mathcal{R}}_{i,k,l}\right). \tag{A.107}$$

Substituting Eq.[A.107] into Eq.[A.105] gives

$$\begin{aligned}\Omega_T \propto & \int \mathrm{d}\mathbf{r}^N \int \mathrm{d}\mathbf{p}^N \int \mathrm{d}\boldsymbol{\mathcal{R}}^{NLM} \int \mathrm{d}\mathbf{p}^{NLM} \\ & \times \exp\left(\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\boldsymbol{\mathcal{R}}_{i,k,l}\right) \\ & \times\delta\left[\sum_{i=1}^{N}\left\{\frac{\mathbf{p}_i^2}{2m_i}+\sum_{k=1}^{M}\sum_{l=1}^{L}\left(\frac{\mathbf{p}_{i,k,l}^2}{2\mathcal{Q}_{i,k,l}}\right)+k_BT\boldsymbol{\mathcal{R}}_{i,k,l}\right\}\right. \\ & \left.+\phi(\mathbf{r}^N)-C\right]. \end{aligned} \tag{A.108}$$

By integrating the δ function that comes from the conserved quantities in the same manner as previous cases would yield

$$\begin{aligned}\Omega_T \propto & \frac{e^{\beta E}}{k_BT}\int \mathrm{d}\mathbf{r}^N \int \mathrm{d}\mathbf{p}^N \int \mathrm{d}\boldsymbol{\mathcal{R}}^{NLM-1} \int \mathrm{d}\mathbf{p}^{NLM} \\ & \times \exp\left[-\beta\left(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}+\phi(\mathbf{r}^N)+\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{L}\frac{\mathbf{p}_{i,k,l}^2}{2\mathcal{Q}_{i,k,l}}\right)\right] \\ & \propto \int \mathrm{d}\mathbf{r}^N \int \mathrm{d}\mathbf{p}^N \exp\left[-\beta\left(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}+\phi(\mathbf{r}^N)\right)\right] \\ & \propto \Omega_{\mathrm{NVT}}\,, \end{aligned} \tag{A.109}$$

where $\beta = 1/kT$.

A.4 Barostating

This section discusses the modification of the extended system previously introduced by Nosé-Hoover to enable one simulating constant pressure system. The scheme was proposed by Shinoda, et al. [78]. The discussion starts from the modifying form of Nosé-Hoover scheme proposed by Martyna et al. [79]. The scheme introduce the additional variable in the same manner as coupling chain of thermostats to the system. Rewriting the equation

of motion in the non-Hamiltonian system by introducing the cell's dynamics is as follow

$$
\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{\mathbf{p}_h}{\mathcal{W}_h}\mathbf{r}_i\,, \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{\mathbf{p}_h}{\mathcal{W}_h}\mathbf{p}_i - \frac{1}{N_f}\frac{\mathrm{Tr}[\mathbf{p}_h]}{\mathcal{W}_h}\mathbf{p}_i - \frac{p_{\mathcal{R}}}{\mathcal{Q}}\mathbf{p}_i\,, \\
\dot{\mathbf{h}} &= \frac{\mathbf{p}_h}{\mathcal{W}_h}\mathbf{h}\,, \\
\dot{\mathbf{p}}_h &= V(\mathbf{P}_{\mathrm{int}} - \mathbf{I}P_{\mathrm{ext}}) - \mathbf{h}\boldsymbol{\xi}\mathbf{h}^t + \left(\frac{1}{N_f}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\right)\mathbf{I} - \frac{p_{\mathcal{R}_1}}{\mathcal{Q}_1}\mathbf{p}_h, \\
\dot{\mathcal{R}}_k &= \frac{p_{\mathcal{R}_k}}{\mathcal{Q}_k} \quad \text{for } k = 1, ..., M, \\
\dot{p}_{\mathcal{R}_1} &= \sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} + \frac{1}{\mathcal{W}_h}\mathrm{Tr}[\mathbf{p}_h^t\mathbf{p}_h] - (N_f + d^2)k_B T_{\mathrm{ext}} - p_{\mathcal{R}_1}\frac{p_{\mathcal{R}_2}}{\mathcal{Q}_2}, \\
\dot{p}_{\mathcal{R}_k} &= \left(\frac{p_{\mathcal{R}_{k-1}}^2}{\mathcal{Q}_{k-1}} - k_B T_{\mathrm{ext}}\right) - p_{\mathcal{R}_k}\frac{p_{\mathcal{R}_{k+1}}}{\mathcal{Q}_{k+1}} \quad \text{for} \quad k = 2, .., M-1 \\
\dot{p}_{\mathcal{R}_M} &= \left(\frac{p_{\mathcal{R}_{M-1}}^2}{\mathcal{Q}_{M-1}} - k_B T_{\mathrm{ext}}\right).
\end{aligned}
\tag{A.110}
$$

Additional variable, p_h denotes cell momenta which is modularly invariant. Subscript k is used to denote the attached $k$sth of thermostat variable. Thermostat and barostat masses are introduced and used with the same purposes as previous MNHC to properly control the fluctuations of the system. P_{ext} and T_{ext} are external hydrostatic pressure and temperature, respectively. The pressure and stress employed to the system, $\mathbf{t}$ is then used to calculate the matrix $\boldsymbol{\xi}$ as follow

$$
\boldsymbol{\xi} = \mathbf{h}_0^{-1}(\mathbf{t} - \mathbf{I}P_{\mathrm{ext}})\mathbf{h}_0^{t-1}\,. \tag{A.111}
$$

Variable of $\mathbf{P}_{\mathrm{int}}$ is for internal pressure. The variable is calculated by using the following equation

$$
(P_{\mathrm{int}})_{\alpha\beta} = \frac{1}{V}\left\{\sum_{i=1}^{N}\frac{(\mathbf{p}_i)_\alpha(\mathbf{p}_i)_\beta}{m_i} + (\mathbf{F}_i)_\alpha(\mathbf{r}_i)_\beta - (\boldsymbol{\phi}'\mathbf{h}')_{\alpha\beta}\right\}, \tag{A.112}
$$

$$
(\phi')_{\alpha\beta} = \frac{\partial\phi(\mathbf{r},\mathbf{h})}{\partial(h)_{\alpha\beta}}. \tag{A.113}
$$

The Hamiltonian resulting from EOM of Eq.[A.110] is conserved and can be written as

$$\begin{aligned}\mathcal{H} = &\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + \phi(\mathbf{r},\mathbf{h}) + P_{\text{ext}}\det[\mathbf{h}] + \frac{1}{2\mathcal{W}_h}\text{Tr}[\mathbf{p}_h^t\mathbf{p}_h] \\ &+\sum_{k=1}^{M}\frac{p_{\mathcal{R}_k}^2}{2\mathcal{Q}_k} + \frac{1}{2}\text{Tr}[\mathbf{h}_0^{-1}(\mathbf{t}-\mathbf{I}P_{\text{ext}})\mathbf{h}_0^{t-1}\boldsymbol{\kappa}] \\ &+(N_f+d^2)k_BT_{\text{ext}}\mathcal{R}_1 + k_BT_{\text{ext}}\sum_{k=2}^{M}\mathcal{R}_k. \end{aligned} \tag{A.114}$$

Here $\boldsymbol{\kappa} = \mathbf{h}^t\mathbf{h}$ is used to denote the metric tensor. The energy contribution comes from the applied external stress, namely elastic energy, can be seen in the sixth right hand side part of Eq.[A.114]. In the same manner from the previous derivation of thermostat scheme, the Jacobian of the coordinate transformation can be written as

$$J = \det[\mathbf{h}]^{1-d}\exp\left[(N_f+d^2)\mathcal{R}_1 + \sum_{k=2}^{M}\mathcal{R}_k\right], \tag{A.115}$$

and substituted to the partition function equation to give

$$\Omega_{\text{P,T}} \propto \int d\mathbf{h}\exp[-\beta\mathcal{H}'']\det[\mathbf{h}]^{1-d}, \tag{A.116}$$

where

$$\begin{aligned}\mathcal{H}'' = &\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + \frac{1}{2\mathcal{W}_h}\text{Tr}[\mathbf{p}_h^t\mathbf{p}_h] + \sum_{k=1}^{M}\frac{p_{\mathcal{R}_k}^2}{2\mathcal{Q}_k} + \phi(\mathbf{r},\mathbf{h}) \\ &+P_{\text{ext}}\det[\mathbf{h}] + \frac{1}{2}\text{Tr}[\mathbf{h}_0^{-1}(\mathbf{t}-\mathbf{E}P_{\text{ext}})\mathbf{h}_0^{t-1}\mathbf{G}]. \end{aligned} \tag{A.117}$$

When the method is applied to the MNHC, the off diagonal cell momenta and kinetic energy should be driven to $k_BT_{\text{ext}}/2$.

A.5 Liouville formulation for Time Reversible Integrator

To integrate equation of motions derived from the previous section, Liouville operator is used in the formalism to satisfy the time-reversible and area-preserving criteria in which suitable for MD simulation. This algorithm firstly proposed by Tuckerman et al. [75]. The method is delivered by the formulation of the equation of motion in a form that is satisfying the Verlet algorithm. This fulfillment can be viewed as the special case of applying Trotter expansion when splitting the energy contributions in writing the system's Hamiltonian. Suppose that f is a function of time dependent variables of $\boldsymbol{\zeta}(t)$. Time derivative of this

function can be written as

$$\begin{aligned} \dot{f} &= \dot{\boldsymbol{\zeta}} \cdot \boldsymbol{\nabla} f \\ &\equiv i\hat{L}_{\boldsymbol{\zeta}} f \,, \end{aligned} \tag{A.118}$$

where

$$i\hat{L}_{\boldsymbol{\zeta}} = \dot{\boldsymbol{\zeta}} \cdot \boldsymbol{\nabla} \,, \tag{A.119}$$

is a Liouville operator of variable $\boldsymbol{\zeta}$. The solution of time integration of Eq.[A.118] can be written as

$$f(\boldsymbol{\zeta}(t)) = \mathrm{e}^{i\hat{L}_{\boldsymbol{\zeta}} t} f(\boldsymbol{\zeta}(t_0)) \,. \tag{A.120}$$

To be more specific, when one need to apply the formulation in a molecular dynamic simulation, the following discussion will describe how the Verlet scheme can be satisfied by applying this Liouville operator. First of all, suppose that the system under condiseration consist of N particles that can be described by position $\mathbf{r}^N$ and momentum $\mathbf{p}^N$. With no external force applied to the system, the Hamiltonian is a function of position $\mathbf{r}^N$ and momentum $\mathbf{p}^N$, namely $\mathcal{H}(\mathbf{r}, \mathbf{p})$ that can be written as

$$\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \phi(\mathbf{r}^N) \,. \tag{A.121}$$

The system's equation of motion can be given by

$$\dot{\mathbf{r}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} \quad \text{and} \quad \dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}_i} = \mathbf{F}_i, \tag{A.122}$$

where Liouville operator is

$$\begin{aligned} i\hat{L} &= \sum_{i=1}^{N} \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} \cdot \frac{\partial}{\partial \mathbf{r}_i} - \frac{\partial \mathcal{H}}{\partial \mathbf{r}_i} \cdot \frac{\partial}{\partial \mathbf{p}_i} \right] \\ &= \sum_{i=1}^{N} \left[\frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{r}_i} + \mathbf{F}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} \right] \\ &\equiv i\hat{L}_{\mathbf{r}} + i\hat{L}_{\mathbf{p}} \,. \end{aligned} \tag{A.123}$$

For non commuting operators such as the above $\mathbf{r}$ and $\mathbf{p}$, the following relation,

$$\exp\left(i\hat{L}_{\mathbf{r}} t + \hat{L}_{\mathbf{p}} t\right) \neq \exp\left(i\hat{L}_{\mathbf{r}} t\right) \exp\left(i\hat{L}_{\mathbf{p}} t\right) \,, \tag{A.124}$$

holds. By introducing the Trotter identity, Eq.A.124 can be rewrite as

$$\exp\left(i\hat{L}_{\mathbf{r}}t+\hat{L}_{\mathbf{p}}t\right)=\lim_{P\to\infty}\left[\exp\left(i\hat{L}_{\mathbf{p}}\frac{t}{2P}\right)\exp\left(i\hat{L}_{\mathbf{r}}\frac{t}{P}\right)\exp\left(i\hat{L}_{\mathbf{p}}\frac{t}{2P}\right)\right]^{P}. \tag{A.125}$$

However, to be used in practical scheme, smaller finite value of P is considered in such a way that the following relation,

$$\exp\left(i\hat{L}_{\mathbf{r}}t+\hat{L}_{\mathbf{p}}t\right)=\left[\exp\left(i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\mathbf{r}}\Delta t\right)\exp\left(i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}\right)\right]^{P}\exp\left[\mathcal{O}\left(\frac{1}{P^2}\right)\right], \tag{A.126}$$

holds; here, $\Delta t=\frac{t}{P}$. The solution of f is a function of those two time-derivative variables. This solution can be obtained by rewriting Eq.[A.120] with the imposing of Trotter expansion as follow

$$\begin{aligned}
f(\mathbf{r}_i(\Delta t),\mathbf{p}_i(\Delta t)) &= \mathrm{e}^{i\hat{L}_{\mathbf{r}}\Delta t+i\hat{L}_{\mathbf{p}}\Delta t}f(\mathbf{r}_i(0),\mathbf{p}_i(0))\\
&= \mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\mathrm{e}^{i\hat{L}_{\mathbf{r}}\Delta t}\mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\begin{pmatrix}\mathbf{r}_i(0)\\ \mathbf{p}_i(0)\end{pmatrix}\\
&= \mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\mathrm{e}^{i\hat{L}_{\mathbf{r}}\Delta t}\left(1+\frac{\Delta t}{2}\mathbf{F}_i(0)\frac{\partial}{\partial\mathbf{p}_i}\right)\begin{pmatrix}\mathbf{r}_i(0)\\ \mathbf{p}_i(0)\end{pmatrix}\\
&= \mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\mathrm{e}^{i\hat{L}_{\mathbf{r}}\Delta t}\begin{pmatrix}\mathbf{r}_i(0)\\ \mathbf{p}_i(\Delta t/2)\end{pmatrix} \qquad \text{(A.127)}\\
&= \mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\left(1+\Delta t\frac{\mathbf{p}_i(\Delta t/2)}{m_i}\frac{\partial}{\partial\mathbf{r}_i}\right)\begin{pmatrix}\mathbf{r}_i(0)\\ \mathbf{p}_i(\Delta t/2)\end{pmatrix}\\
&= \mathrm{e}^{i\hat{L}_{\mathbf{p}}\frac{\Delta t}{2}}\begin{pmatrix}\mathbf{r}_i(\Delta t)\\ \mathbf{p}_i(\Delta t/2)\end{pmatrix} \qquad \text{(A.128)}\\
&= \left(1+\frac{\Delta t}{2}\mathbf{F}_i(\Delta t)\frac{\partial}{\partial\mathbf{p}_i}\right)\begin{pmatrix}\mathbf{r}_i(\Delta t)\\ \mathbf{p}_i(\Delta t/2)\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{r}_i(\Delta t)\\ \mathbf{p}_i(\Delta t)\end{pmatrix}. \qquad \text{(A.129)}
\end{aligned}$$

The introduction of Trotter identity to the discretized version of formal Liouville equation have resulted an integration scheme similar to those seen in the velocity Verlet algorithm;

here, the time propagation of $\mathbf{r}(t)$ and $\mathbf{p}(t)$ can be described as

$$\begin{aligned}\mathbf{p}_i(\Delta t/2) &\leftarrow \mathbf{p}_i(0) + \frac{\Delta t}{2}\,\mathbf{F}_i(0)\\ \mathbf{r}_i(\Delta t) &\leftarrow \mathbf{r}_i(0) + \Delta t\,\frac{\mathbf{p}_i(\Delta t/2)}{m_i}\\ \mathbf{p}_i(\Delta t) &\leftarrow \mathbf{r}_i(\Delta t) + \mathbf{F}_i(\Delta t)\frac{\Delta t}{2}\,.\end{aligned} \tag{A.130}$$

Suppose that the system under study is subjected to Thermostat and Barostat where the equation of motion follows Eq.[A.110]. Several factorization schemes can be used to formulate the time integration. The rRESPA algorithm is used as the basis to generate explicit factorization of the time reversible integration scheme of EOM. According to Shinoda et al. [78], Liouville formulation of variable's time propagation of their original EOM form is written as follow

$$\begin{aligned}\exp(i\hat{L}\Delta t) = \exp\left(i\hat{L}_{\mathrm{bath}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_2\frac{\Delta t}{2}\right)\exp(i\hat{L}_1\Delta t)\\ \times\exp\left(i\hat{L}_2\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\mathrm{bath}}\frac{\Delta t}{2}\right) + \mathcal{O}(\Delta t^3)\,.\end{aligned} \tag{A.131}$$

The Liouville operators are given below

$$i\hat{L} = i\hat{L}_1 + i\hat{L}_2 + i\hat{L}_{\mathrm{bath}}\,, \tag{A.132}$$

with

$$i\hat{L}_1 = \sum_{i=1}^{N}[\mathbf{v}_i + \mathbf{v}_h\mathbf{r}_i]\cdot\boldsymbol{\nabla}_{\mathbf{r}_i} + \sum_{\alpha,\beta}(\mathbf{v}_h)_{\alpha\beta}\frac{\partial}{\partial(\mathbf{h})_{\alpha\beta}}, \tag{A.133}$$

$$i\hat{L}_2 = \sum_{i=1}^{N}\left[\frac{\mathbf{F}_i}{m_i}\right]\cdot\boldsymbol{\nabla}_{\mathbf{v}_i}, \tag{A.134}$$

$$
\begin{aligned}
i\hat{L}_{\text{bath}} = \sum_{i=1}^{N}\Bigg[& -\left\{\mathbf{v}_h + \left(\frac{1}{N_f}\right)\text{Tr}[\mathbf{v}_h] + v_{\mathcal{R}_1}\right\}\mathbf{v}_i\Bigg]\cdot\boldsymbol{\nabla}_{\mathbf{v}_i} \\
& + \sum_{\alpha,\beta}\Bigg\{\frac{1}{\mathcal{W}_h}\Bigg[\sum_{i=1}^{N} m_i(\mathbf{v}_i)_\alpha(\mathbf{v}_i)_\beta + \sum_{i=1}^{N}(\mathbf{F}_i)_\alpha(\mathbf{r}_i)_\beta - (\boldsymbol{\phi}'\mathbf{h}^t)_{\alpha\beta} \\
& + \left(\frac{1}{N_f}\sum_{i=1}^{N} m_i\mathbf{v}_i^2 - P_{\text{ext}}V\right)\delta_{\alpha\beta} - [\mathbf{h}\mathbf{h}_0^{-1}(\mathbf{t} - \mathbf{I}P_{\text{ext}})\mathbf{h}_0^{t-1}\mathbf{h}^t]_{\alpha\beta}\Bigg] \\
& - v_{\mathcal{R}_1}(\mathbf{v}_h)_{\alpha\beta}\Bigg\}\frac{\partial}{\partial(\mathbf{v}_h)_{\alpha\beta}} + \sum_{k=1}^{M} v_{\mathcal{R}_k}\frac{\partial}{\partial\mathcal{R}_k} \\
& + \left[\frac{1}{\mathcal{Q}_1}\left(\sum_{i=1}^{N} m_i\mathbf{v}_i^2 + \mathcal{W}_h\text{Tr}[\mathbf{v}_h^t\mathbf{v}_h] - (N_f + d^2)kT_{\text{ext}}\right) - v_{\mathcal{R}_1}v_{\mathcal{R}_2}\right]\frac{\partial}{\partial v_{\mathcal{R}_1}} \\
& + \sum_{k=2}^{M-1}\left[\frac{1}{\mathcal{Q}_k}(\mathcal{Q}_{k-1}v_{\mathcal{R}_{k-1}}^2 - kT_{\text{ext}}) - v_{\mathcal{R}_k}v_{\mathcal{R}_{k+1}}\right]\frac{\partial}{\partial v_{\mathcal{R}_k}} \\
& + \left[\frac{1}{\mathcal{Q}_M}(\mathcal{Q}_{M-1}v_{\mathcal{R}_{M-1}}^2 - kT_{\text{ext}})\right]\frac{\partial}{\partial v_{\mathcal{R}_M}}
\end{aligned}
\tag{A.135}
$$

where $\mathbf{v}_i = \mathbf{p}_i/m_i \neq \dot{\mathbf{r}}_i$, $\mathbf{v}_h = \mathbf{p}_h/\mathcal{W}_h$, and $v_{\mathcal{R}_k} = p_{\mathcal{R}_k}/\mathcal{Q}_k$.

Following Tuckerman et al. [75], the factorization to generate explicit reversible time integrator for the NPT ensemble previously used in Martyna et al. [79] is now can be written as follow

$$
\begin{aligned}
\exp(i\hat{L}\Delta t) = & \exp\left(i\hat{L}_{\text{T}-\text{baro}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\text{T-part}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\varepsilon,2}^{(2)}\frac{\Delta t}{2}\right) \\
& \times\exp\left(i\hat{L}_2^{(2)}\frac{\Delta t}{2}\right)\Bigg[\exp\left(i\hat{L}_{\varepsilon,2}^{(1)}\frac{\delta t}{2}\right)\exp\left(i\hat{L}_2^{(1)}\frac{\delta t}{2}\right) \\
& \times\exp\left(i\hat{L}_{\varepsilon,1}\delta t\right)\exp\left(i\hat{L}_1\delta t\right)\exp\left(i\hat{L}_2^{(1)}\frac{\delta t}{2}\right)\exp\left(i\hat{L}_{\varepsilon,2}^{(1)}\frac{\delta t}{2}\right)\Bigg]^n \\
& \times\exp\left(i\hat{L}_2^{(2)}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\varepsilon,2}^{(2)}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\text{T-part}}\frac{\Delta t}{2}\right) \\
& \times\exp\left(i\hat{L}_{\text{T}-\text{baro}}\frac{\Delta t}{2}\right) + \mathcal{O}(\Delta t^3)\,.
\end{aligned}
\tag{A.136}
$$

The scheme used two time steps, δt and $\Delta t = n\delta t$. Within this scheme, the Liouville formalization is written as

$$
i\hat{L}_1 = \sum_{i=1}^{N}[\mathbf{v}_i + \mathbf{v}_h\mathbf{r}_i]\cdot\boldsymbol{\nabla}_{\mathbf{r}_i}\,. \tag{A.137}
$$

For the fast operator at $i\hat{L}_2$, the scheme gives

$$
i\hat{L}_2^{(1)} = \sum_{i=1}^{N}\left[\left[\frac{\mathbf{F}_i}{m_i}\right]^{(1)} - \left\{\mathbf{v}_h + \frac{\text{Tr}[\mathbf{v}_h]}{N_f}\right\}\mathbf{v}_i\right]\cdot\boldsymbol{\nabla}_{\mathbf{v}_i}\,, \tag{A.138}
$$

On the other hand, the slower operator of $i\hat{L}_2$ can be written as

$$i\hat{L}_2^{(2)} = \sum_{i=1}^{N} \left[\frac{\mathbf{F}_i}{m_i}\right]^{(2)} \cdot \boldsymbol{\nabla}_{\mathbf{v}_i} . \tag{A.139}$$

For ε of the barostat variable explained in [75], Liouville operator can be given as

$$\begin{aligned} i\hat{L}_{\varepsilon,1} &= \sum_{\alpha,\beta} (\mathbf{v}_h)_{\alpha\beta} \frac{\partial}{\partial(\mathbf{h})_{\alpha\beta}} , \\ i\hat{L}_{\varepsilon,2}^{(1)} &= G_{\varepsilon}^{(1)} \frac{\partial}{\partial(\mathbf{v}_h)_{\alpha\beta}} , \\ i\hat{L}_{\varepsilon,2}^{(2)} &= G_{\varepsilon}^{(2)} \frac{\partial}{\partial(\mathbf{v}_h)_{\alpha\beta}} . \end{aligned} \tag{A.140}$$

Translation's operator of the thermostat attached to the barostat is given by

$$i\hat{L}_{\text{T-baro}} = \frac{1}{\mathcal{Q}_1} \mathcal{W}_g \text{Tr}[\mathbf{v}_h^t \mathbf{v}_h] \frac{\partial}{\partial v_{\mathcal{R}_1}} - \sum_{\alpha,\beta} v_{\mathcal{R}_1} (\mathbf{v}_h)_{\alpha\beta} \frac{\partial}{\partial(\mathbf{v}_h)_{\alpha\beta}} , \tag{A.141}$$

where that attached to the system's particles is given by

$$\begin{aligned} i\hat{L}_{\text{T-par}} = &- \sum_{i=1}^{N} v_{\mathcal{R}_1} \mathbf{v}_i \cdot \boldsymbol{\nabla}_{\mathbf{v}_i} + \sum_{k=1}^{M} v_{\mathcal{R}_k} \frac{\partial}{\partial \mathcal{R}_k} , \\ &- \left\{ \frac{(N_f + d^2) k T_{\text{ext}}}{\mathcal{Q}_1} + v_{\mathcal{R}_1} v_{\mathcal{R}_2} \right\} \frac{\partial}{\partial v_{\mathcal{R}_1}} , \\ &+ \sum_{k=2}^{M-1} \left[\frac{1}{\mathcal{Q}_k} (\mathcal{Q}_{k-1} v_{\mathcal{R}_{k-1}}^2 - k T_{\text{ext}}) - v_{\mathcal{R}_k} v_{\mathcal{R}_{k+1}} \right] \frac{\partial}{\partial v_{\mathcal{R}_k}} , \\ &+ \left[\frac{1}{\mathcal{Q}_M} (\mathcal{Q}_{M-1} v_{\mathcal{R}_{M-1}^2} - k T_{\text{ext}}) \right] \frac{\partial}{\partial v_{\mathcal{R}_M}} , \end{aligned} \tag{A.142}$$

where

$$\begin{aligned} G_{\varepsilon}^{(1)} = \frac{1}{\mathcal{W}_h} \sum_{\alpha,\beta} \Bigg\{ &\sum_{i=1}^{N} \left[\left\{ \frac{m_i \mathbf{v}_i^2}{N_f} + -P_{\text{ext}}^{(1)} V \right\} \delta_{\alpha\beta} + m_i (\mathbf{v}_i)_\alpha (\mathbf{v}_i)_\beta + (\mathbf{F}_i)_\alpha (\mathbf{r}_i)_\beta \right] \\ &- [\mathbf{h}\mathbf{h}_0^{-1} (\mathbf{t} - \mathbf{I} P_{\text{ext}}^{(1)}) \mathbf{h}_0^{t-1} \mathbf{h}^t]_{\alpha\beta} - \frac{\partial \phi(\mathbf{r}, \mathbf{h})}{\partial (h)_{\alpha\beta}} \mathbf{h}_{\alpha\beta}^t \Bigg\} , \end{aligned} \tag{A.143}$$

and

$$G_{\varepsilon}^{(2)} = \frac{1}{\mathcal{W}_h}\sum_{\alpha,\beta}\left\{\sum_{i=1}^{N}\left[(\mathbf{F}_i)_\alpha(\mathbf{r}_i)_\beta - P_{\text{ext}}^{(2)}V\delta_{\alpha\beta}\right]\right.$$
$$\left.-[\mathbf{h}\mathbf{h}_0^{-1}(\mathbf{t}-\mathbf{I}P_{\text{ext}}^{(2)})\mathbf{h}_0^{t-1}\mathbf{h}^t]_{\alpha\beta} - \frac{\partial\phi(\mathbf{r},\mathbf{h})}{\partial(h)_{\alpha\beta}}\mathbf{h}_{\alpha\beta}^t\right\}. \tag{A.144}$$

Superscripts 1 and 2 are given to label the fast and the slow contributions, respectively. This scheme is employed in LAMMPS "fix npt" style to integrate the equation of motions of NPT ensemble.

For the practical purposes, the factorization of the Liouville formulation can be done in a slightly different form. The view is based on the condition that the barostat only updated in the outward side of the rRESPA level. Numerical stability is guaranteed in Tuckerman et al. [75] scheme. However, when damping factor is employed to determine the time scale of pressure's relaxation with the amplitude of several outermost parts of the rRESPA, numerical stability is no longer correspond to the barostat dynamics. Rewriting the factorization scheme with this slight modification can be given as

$$\begin{aligned}\exp(i\hat{L}\Delta t) = {} & \exp\left(i\hat{L}_{\text{T}-\text{baro}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\text{T-part}}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\varepsilon,2}\frac{\Delta t}{2}\right)\\ & \times\exp\left(i\hat{L}_2^{(2)}\frac{\Delta t}{2}\right)\left[\exp\left(i\hat{L}_2^{(1)}\frac{\Delta t}{2n}\right)\exp\left(i\hat{L}_{\varepsilon,1}\frac{\Delta t}{2n}\right)\right.\\ & \left.\times\exp\left(i\hat{L}_1\frac{\Delta t}{n}\right)\exp\left(i\hat{L}_{\varepsilon,1}\frac{\Delta t}{2n}\right)\exp\left(i\hat{L}_2^{(1)}\frac{\Delta t}{2n}\right)\right]^n\\ & \times\exp\left(i\hat{L}_2^{(2)}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\varepsilon,2}\frac{\Delta t}{2}\right)\exp\left(i\hat{L}_{\text{T-part}}\frac{\Delta t}{2}\right)\\ & \times\exp\left(i\hat{L}_{\text{T}-\text{baro}}\frac{\Delta t}{2}\right)+\mathcal{O}(\Delta t^3)\,.\end{aligned} \tag{A.145}$$

The scheme is used in LAMMPS "fix npt" style with "respa" option in "fixes style".

A.6 Constraints

This section will discuss the use of constraints to obtain the constant bond length and angle of atom pairs concerning the integration time. The constraint is very useful when one needs to simulate rigid molecule models such as provided in water molecule and protein. Therefore, the applied method should be suitable and easy to implement in the well-known integration scheme.

The discussion starts with the simplest implementation of constraints in molecular dynamics simulation. The constraint was implemented by using Lagrange multiplier-

based method. Suppose that there are l linear-holonomic constraints at the time t in N interacting particle system,

$$\sigma_\beta(t) \equiv (\mathbf{r}_{\beta j}(t) - \mathbf{r}_{\beta i}(t))^2 - d_\beta^2 = 0 \qquad \beta = 1, 2, ..., l \tag{A.146}$$

Subscript β is used to indicate constraints while i and j are used for the particle. Equation of motion for the introduction of holonomic constraints in the system such as Eq.[A.146] is given by

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i + \mathbf{G}_i = -\boldsymbol{\nabla}_i \phi(\{\mathbf{r}\}) - \sum_{k=1}^{l} \lambda_k \boldsymbol{\nabla}_i \sigma_k \tag{A.147}$$

where $\mathbf{G}_i$ is the force on particle i generated by all constraints correspond to i, and λ_k is Lagrange multiplier of σ_k constraint. Integrating both sides of Eq.[A.147] gives

$$\mathbf{r}_i(t + \Delta t) = \acute{\mathbf{r}}_i(t + \Delta t) + \frac{(\Delta t)^2}{m_i} \sum_{k=1}^{l} \lambda_k \boldsymbol{\nabla}_i \sigma_k(t)\,, \tag{A.148}$$

where $\acute{\mathbf{r}}(t + \Delta t)$ is particle's position vector without constraints. The constraints of next Δt time step can be given by

$$\sigma_\beta(t + \Delta t) \equiv (\mathbf{r}_{\beta j}(t + \Delta t) - \mathbf{r}_{\beta i}(t + \Delta t))^2 - d_\beta^2 = 0\,. \tag{A.149}$$

By substituting Eq.[A.148] to Eq.[A.149], constraints of the next Δt step are given as follow

$$\begin{aligned} \sigma_\beta(t + \Delta t) \equiv \Bigg\{ & \acute{\mathbf{r}}_{\beta j}(t + \Delta t) - \acute{\mathbf{r}}_{\beta i}(t + \Delta t) \\ & + (\Delta t)^2 \sum_{k=1}^{l} \lambda_k \left(\frac{\boldsymbol{\nabla}_{\beta j} \sigma_k(t)}{m_{\beta j}} - \frac{\boldsymbol{\nabla}_{\beta i} \sigma_k(t)}{m_{\beta i}} \right) \Bigg\}^2 - d_\beta^2 = 0\,, \end{aligned} \tag{A.150}$$

which is l simultaneous non-linear equation. To solve this, Newton-Raphson method of

$$\boldsymbol{\lambda}^{(\gamma+1)} \leftarrow \boldsymbol{\lambda}^{(\gamma)} - \mathbf{J}_\sigma^{-1} \boldsymbol{\sigma}(t + \Delta t) \tag{A.151}$$

can be applied. The Jacobian, $\mathbf{J}$ written in the above equation is given by

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \sigma_1}{\partial \lambda_1} & \frac{\partial \sigma_1}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_1}{\partial \lambda_l} \\ \frac{\partial \sigma_2}{\partial \lambda_1} & \frac{\partial \sigma_2}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_2}{\partial \lambda_l} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_l}{\partial \lambda_1} & \frac{\partial \sigma_l}{\partial \lambda_2} & \cdots & \frac{\partial \sigma_l}{\partial \lambda_l} \end{pmatrix}. \tag{A.152}$$

This Jacobian can be solved on each molecule under constraint. This glimpse is based on

the fact that the contribution to the constraints comes only from a specific particles in the system. It is easier and more simpler way to start with $\boldsymbol{\lambda}^{(0)} = 0$ to give

$$J_{\alpha\beta} = \left.\frac{\partial \sigma_\beta}{\partial \lambda_\alpha}\right|_{\lambda=0} = 2\left(\mathbf{r}_{\beta j} - \mathbf{r}_{\beta i}\right)\left\{\frac{\partial \sigma_\alpha}{\partial \mathbf{r}_{\beta j}} m_{\beta j}^{-1} - \frac{\partial \sigma_\alpha}{\partial \mathbf{r}_{\beta i}} m_{\beta i}^{-1}\right\}, \tag{A.153}$$

and solving for the updated λ to be written as

$$\lambda_\beta = -\mathbf{J}^{-1}\left[\left(\mathbf{r}_{\beta j}(t+\Delta t) - \mathbf{r}_{\beta i}(t+\Delta t)\right)^2 - d_\beta^2\right]. \tag{A.154}$$

Position vectors on each iteration are updated by using

$$\mathbf{r}_i(t+\Delta t) \leftarrow \mathbf{r}_i(t+\Delta t) + \sum_{k=1}^{l} \lambda_k \frac{\partial \sigma_k}{\partial \mathbf{r}_i}. \tag{A.155}$$

In order to start the new iteration to achieve a convergence of $\sigma_(t+\Delta t)$ to a given tolerance value, $\boldsymbol{\lambda}$ should be reset to 0.

A.6.1 SHAKE algorithm

This algorithm uses the same Lagrange multiplier-based method to incorporate constraint in the molecular dynamics simulation. It intended to be properly applied in Verlet integration scheme by employing a set of $\{\alpha_\beta\}$ in the calculation to meet the integrator accuracy, thus avoiding the need to calculate the derivation of λ up to order $n-2$. A slightly different scheme of establishing the solution of Lagrange multiplier through the analytical one is presented. Firstly, consider Lagrange multiplier parameter, λ as written below

$$\lambda_k(t) = \sum_{n=0}^{\infty} \lambda_k^{(n)}(t_0) \frac{(t-t_0)^n}{n!}. \tag{A.156}$$

Suppose that a system of N interacting particles is subjected to a constraint declared in Eq.[A.146]. The taylor expansion of particle's position vector at time t from t_0 with the introduction of Lagrange multiplier (written in Eq.[A.156]) can be given as

$$\begin{aligned}
\mathbf{r}_i(t, \{\lambda()\}) &\equiv \mathbf{r}(t, \{\lambda^{(n)}(t_0)\}) \\
&= \mathbf{r}_i(t_0) + \dot{\mathbf{r}}_i(t_0)(t-t_0) + \frac{1}{m_i}\sum_{n=2}^{\infty} \frac{(t-t_0)^n}{n!} \cdot \left\{ \mathbf{F}_i^{(n-2)}(t_0) \right. \\
&\quad \left. - \sum_{k=1}^{l}\sum_{p=0}^{n-2} \binom{n-2}{p} \lambda_k^{(p)}(t_0) \left[\left(\boldsymbol{\nabla}_i \sigma_k\right)^{(n-2-p)}\right]_{t_0} \right\},
\end{aligned} \tag{A.157}$$

where force derivation $\mathbf{F}_i^{(s)}$ is given by

$$\mathbf{F}_i^{(s)}(t_0) \equiv \left[\frac{d^s}{dt^2}\mathbf{F}_i(t_0)\right] = \sum_{j=1}^{N} \frac{\partial \mathbf{F}_i^{(s-1)}(t_0)}{\partial \mathbf{r}_j} \cdot \dot{\mathbf{r}}_j(t_0)\,, \tag{A.158}$$

and the time derivative of $\mathbf{G}_i$, which is a force generated by constraints, is written as follow

$$\left[(\boldsymbol{\nabla}_i \sigma_\beta)^{(s)}\right]_{t=t_0} \equiv \left[\frac{d^2}{dt^s}(\boldsymbol{\nabla}_i \sigma_\beta)\right]_{t=t_0} = \sum_{j=1}^{N} \boldsymbol{\nabla}_j \left[(\boldsymbol{\nabla}_i \sigma_\beta)^{(s-1)}\right]_{t_0} \cdot \dot{\mathbf{r}}_j(t_0)\,. \tag{A.159}$$

To get the solution of $\lambda_\beta^{(s)}$ set for $s = 0, 1, \ldots$, consider the constraints equation of Eq.[A.156]. Regarding the information contained in the equation, the time derivative will be equal to 0 since the constraints are hold for all of the simulation time. This condition gives

$$\sigma_\beta^{(s+2)}(\{\mathbf{r}(t_0)\}) \equiv \left[\frac{d^{s+2}}{dt^{s+2}}\sigma_\beta(\{\mathbf{r}(t)\})\right]_{t_0} = 0 \qquad (\beta = 1, 2, \ldots, l)\,, \tag{A.160}$$

with the explicit derivation is the following

$$\begin{aligned}
\sigma_\beta^{(s+2)}(\{\mathbf{r}(t_0)\}) &= \sum_{i=1}^{N}\sum_{\gamma=0}^{s+1} \binom{s+1}{\gamma} \mathbf{r}_i^{(s+2-\gamma)}(t_0) \cdot \left[(\boldsymbol{\nabla}_i \sigma_\beta)^{(\gamma)}\right]_{t_0} \\
&= \sum_{i=1}^{N}\sum_{\gamma=0}^{s+1} \binom{s+1}{\gamma} \frac{1}{m_i} \Bigg\{ \mathbf{F}_i^{(s-\gamma)}(t_0) \\
&\quad - \sum_{p=1}^{l}\sum_{\kappa=0}^{s-\gamma} \binom{s-\gamma}{\kappa} \lambda_p^{(\kappa)}(t_0) \left[(\boldsymbol{\nabla}_i \sigma_p)^{(s-\gamma-\kappa)}\right]_{t_0} \Bigg\} \left[(\boldsymbol{\nabla}_i \sigma_k)^{(\gamma)}\right]_{t_0}\,. \\
&= 0
\end{aligned} \tag{A.161}$$

Extracting the contribution of the highest derivative; which is when $\gamma = 0$ and $\kappa = s$; will give

$$\begin{aligned}
&\sum_{i=1}^{N} \frac{1}{m_i} \left\{ \mathbf{F}_i^{(s)}(t_0) - \sum_{p=1}^{l} \lambda_p^{(s)}(t_0) \left[\boldsymbol{\nabla}_i \sigma_p\right]_{t_0} \right\} \cdot \left[\boldsymbol{\nabla}_i \sigma_\beta\right]_{t_0} \\
&+ \mathcal{F}\left(\{\lambda_p^{(s-1)}(t_0), \ldots, \lambda_p^{(0)}(t_0)\}, \{\mathbf{r}(t_0), \dot{\mathbf{r}}(t_0)\}\right) = 0\,,
\end{aligned} \tag{A.162}$$

where the solution of $\lambda_\beta^{(s)}$ set can be obtained by solving the linear equation of Eq.[A.162] iteratively. This stage can be done since $\mathcal{F}$ is a known variable for the time derivation of λ less than s. By doing this and subsequently followed by the substitution of the result

to Eq.[A.147], one will get the ordinary second order differential equations that can be integrated numerically.

However, when this method is closely followed when performing the simulation, additional errors can be generated in a specific integration algorithm as the solution only fulfilling some others. Therefore, probabilities of deviation to proliferate resembling linearly in each time step.

Suppose that the integration algorithm suggesting one to include a fraction of time derivative of space coordinates up to the order of n. Denoting $\mathbf{r}_i^A(t)$ as the values that can be given by the algorithm for $\mathbf{r}_i$, the set of $\lambda^{(0)}(t_0), \ldots, \lambda^{(n-2)}$ have to be calculated to give

$$\sigma\left(\{\mathbf{r}^A(t)\}\right) = \mathcal{O}\left[(\Delta t)^{(m+1)}\right] \tag{A.163}$$

in the order of the worst error that can be made; here, the maximum order of error is comparable to the error order of Taylor expansion when the first order's truncation is employed. Indeed, the n time derivatives of space coordinate require the calculation of time derivative of force up to the order of $n-2$.

It is proven that by replacing $\{\lambda_\beta^{(n-2)}(t_0)\}$ with $\{\alpha_\beta\}$ for which satisfying the relation

$$\sigma_\beta\left(\{\mathbf{r}(t, \acute{\alpha_k})\}\right) = 0\,, \tag{A.164}$$

and calculating only the first $n-3$ derivatives of λ, one can get the same $\mathcal{O}\left[(\Delta t)^{m+1}\right]$ error difference between calculating trajectory by using $\{\alpha_\beta(t_0)\}$ and with $\lambda_\beta^{(n-2)}(t_0)$; error produced by this manner is in the same order of error generated by the implicit algorithm, thus providing new scheme while fulfilling the constraints relation. Furthermore, by employing this new scheme, the equation of trajectory evolution at time Δt can be given as

$$\begin{aligned}\mathbf{r}_i(t_0 + \Delta t, \{\lambda_\beta^{(0)}(t_0), \ldots, \lambda_\beta^{(n-3)}(t_0), \alpha_\beta\}) \\ = \mathbf{r}_i'(t_0 + \Delta t, \{\lambda_\beta^{(0)}(t_0), ..., \lambda_\beta^{(n-3)}(t_0)\}) + \delta\mathbf{r}_i(t_0 + \Delta t, \{\alpha_\beta\})\,.\end{aligned} \tag{A.165}$$

Substituting Eq.[A.165] to Eq.[A.146] gives

$$2(\mathbf{r}_j' - \mathbf{r}_i') \cdot (\delta\mathbf{r}_j - \delta\mathbf{r}_i) + (\delta\mathbf{r}_j - \delta\mathbf{r}_i)^2 = d_{ij}^2 - (\mathbf{r}_j' - \mathbf{r}_i')^2\,, \tag{A.166}$$

where

$$\delta\mathbf{r}_i(t_0 + \Delta t, \{\alpha_\beta\}) = \frac{1}{m_i}\frac{(\Delta t)^n}{n!}\sum_{k=1}^{l}\alpha_\beta(\boldsymbol{\nabla}_i\sigma_\beta)_{t_0}\,. \tag{A.167}$$

Implementing the iteration in solving the trajectory equation of Eq.[A.166] can be initiated by taking $\{\alpha_\beta\} = 0$. This initiation can be done due to the fact that the nonlinear term

in α_β are small when $\Delta t \to 0$.

In the case of Verlet algorithm, it can be seen from Eq.[A.167] that there is no need to calculate λ since n is equal to 2. To be more specific in employing the formulation to the Verlet algorithm, Eq.[A.167] with the introduction of time-step h can be rewrite as

$$\mathbf{r}_i(h) = \mathbf{r}_i(h, \{\alpha_\beta\}) = \mathbf{r}'_i(h) + \delta\mathbf{r}_i(h)\,, \tag{A.168}$$

where

$$\mathbf{r}'_i(h) = -\mathbf{r}_i(-h) + 2\mathbf{r}_i(0) + \left(\frac{h^2}{m_i}\right)\mathbf{F}_i(0) \tag{A.169}$$

denotes the time evolution of particles trajectory with the absence of constraint, and

$$\delta\mathbf{r}_i(h) = \left(\frac{h^2}{m_i}\right)\sum_{k=1}^{l}\alpha_k\left[\boldsymbol{\nabla}_i\sigma_k\right]_{t_0} \tag{A.170}$$

is the correction part generated by the constraint. The set of α_β are resulted from

$$\sigma\left(\{\mathbf{r}_i(h, \{\alpha_\beta\})\}\right) = 0 \quad (\beta = 1, \ldots, l)\,. \tag{A.171}$$

By using Eq.[A.168], Eq.[A.169] and Eq.[A.170] to replace the parameter previously written in Eq.[A.169] gives

$$\begin{aligned} 2(\mathbf{r}'_j(h) - \mathbf{r}'_i(h))\Bigg(& - h^2\sum_{k=1}^{l}\alpha_k\left[\left(\frac{\boldsymbol{\nabla}_j}{m_j} - \frac{\boldsymbol{\nabla}_i}{m_i}\right)\sigma_k\right]_{t=0} \\ & + h^4\sum_{k=1}^{l}\sum_{k'=1}^{l}\alpha_k\alpha'_k\left[\left(\frac{\boldsymbol{\nabla}_j}{m_j} - \frac{\boldsymbol{\nabla}_i}{m_i}\right)\sigma_k\right]_{t=0}\left[\left(\frac{\boldsymbol{\nabla}_j}{m_j} - \frac{\boldsymbol{\nabla}_i}{m_i}\right)\sigma'_k\right]_{t=0}\Bigg) \\ & = d_{ij}^2 - (\mathbf{r}'_j(h) - \mathbf{r}'_i(h))^2\,. \end{aligned} \tag{A.172}$$

This equation can be solved by iteratively update the value of α_β. The value of $\{\alpha_\beta^{[m]}\}$ at M iteration can be obtained by substituting the value of previous $\{\alpha_\beta^{[m-1]}\}$ initiated by $\{\alpha_\beta^{[0]}\} = 0$ in the very beginning of run. Since quadratic form of α_β corresponds to the small h^4 term when one using small time-step, rapid convergence can be easily achieved.

A.6.2 RATTLE algorithm

There is an issue when one is trying to simulate a system with NPT condition subjected to constraints. The need for calculating velocity is explicitly shown in the equation of motion of constant pressure ensemble. The previously described SHAKE algorithm uses Verlet algorithm with velocity is obtained only by employing additional calculation such as finite difference scheme; here, less accuracy and difficulties would accompany the implementation.

Correspond to the problem, several method to modify the SHAKE algorithm were proposed with persistent difficulties in implementation to the recent developed molecular dynamics methods. One would argue that by using other methods of constraint, the result would satisfy the need of velocity value within the integration process when simulating NPT condition. However, instead of the Verlet, the constraint scheme can only work in other integration methods, thus suggesting not suitable to be employed in a system subjected to the thermostat and barostat scheme that is previously described.

RATTLE algorithm used velocity Verlet integration method since it provides explicitly velocity value as a part of the calculation processes. Therefore, RATTLE is suitable to alleviate the problem in simulating atomic fluid and vibrated molecules model; here, the previous SHAKE modified scheme inconsistently worked for the simple iterative algorithm of the velocity VERLET algorithm. This algorithm can be viewed as a general form of SHAKE algorithm in a manner of providing the need of velocity value in constant pressure and temperature system's equation of motion.

Assuming the forces, positions and velocities of time t are already known in the system of N interacting particles subjected to constraints. The particle's position of time $t + h$ can be written as

$$r(t+h) = r(t) + h\dot{r}(t) + \frac{h^2}{2}\left[F[r(t)] + G_R(t)\right] \tag{A.173}$$

where G_R denotes the approximated forces generated by constraints. To get velocity at time $t+h$, by introducing the second approximated force associated with constraints, G_V, time derivative of position at $t + h$ can be given by

$$\dot{r}(t+h) = \dot{r}(t) + \frac{h}{2}\left[F[r(t)] + G_R(t) + F[r(t+h)] + G_V(t)\right]. \tag{A.174}$$

The G_R and G_V are the two separated approximations which are possible to implement to satisfy the constraints shown in Eq.[A.146]. It is worth to mention that this algorithm, as can be seen in Eq.[A.173] and Eq.[A.174], use the one-time calculation of positions and velocities when obtaining the next time-step of positions and velocities, respectively. Furthermore, the calculated positions and velocities at every point in time satisfying the constraints. Moreover, the produced error is the same as the velocity Verlet would produce; it is proven to produce better accuracy compared to the original Verlet algorithm. Besides, the use of initial positions and velocities gives advantages when one need to implement stochastic collision with the magnitude's change during the simulation time.

The algorithm is governed by the following. Suppose that a system consist of N interacting particles subjected to constraints. Similar to the previous SHAKE algorithm, the identical constraints as in Eq.[A.146] were exist in the system under consideration. The time derivative of the constraints as written in the following,

$$[\dot{\mathbf{r}}_i(t) - \dot{\mathbf{r}}_j(t)] \cdot [\mathbf{r}_i(t) - \mathbf{r}_j(t)] = 0 \tag{A.175}$$

should hold for the velocity constraint. Lagrange multiplier-base method is used when governing the RATTLE scheme in the same manner as the SHAKE algorithm. RATTLE equation in Eq.[A.173] correspond to the equation of motion with additional force associated with constraints shown in Eq.[A.147] can be written as

$$\mathbf{r}_i(t+h) = \mathbf{r}_i(t) + h\dot{\mathbf{r}}_i(t) + \frac{h^2}{2m_i}\left[\mathbf{F}_i(t) - 2\sum_{k=1}^{l}\lambda_{Rk}(t)\mathbf{r}_k(t)\right]. \tag{A.176}$$

The chosen $\lambda_{Rk}(t)$ should satisfy the constraints at time $t+h$. Subsequently, the velocity equation of RATTLE can be written as

$$\begin{aligned}\dot{\mathbf{r}}_i(t+h) = \dot{\mathbf{r}}_i(t) + h\dot{\mathbf{r}}_i(t) + \frac{h}{2m_i}\Bigg[&\left(\mathbf{F}_i(t) - 2\sum_{k=1}^{l}\lambda_{Rk}(t)\mathbf{r}_k(t)\right)\\ &\times\left(\mathbf{F}_i(t+h) - 2\sum_{k=1}^{l}\lambda_{Vk}(t+h)\mathbf{r}_k(t+h)\right)\Bigg].\end{aligned} \tag{A.177}$$

The implementation started the picking one of the constraint variable to calculate. In doing this, several variables are propose in the following

$$\begin{aligned}G_\beta &= h\lambda_{R\beta}(t),\\ \gamma_\beta &= h\lambda_{V\beta}(t+h),\\ \mathbf{q}_i &= \dot{\mathbf{r}}_i(t) + \frac{h}{2m_i}\mathbf{F}_i(t) - \frac{1}{m_i}\sum_{k=1}^{l}G_k\mathbf{r}_k(t),\end{aligned} \tag{A.178}$$

in which will be used and calculated iteratively. The chosen constraint are then used to obtain the position's vector of particles. For particle i and j, the vector can written as

$$\mathbf{q}_\kappa = \dot{\mathbf{r}}(t) + \frac{h}{2m_i}\mathbf{F}_i(t) \quad (\kappa = 1, 2, ..., N), \tag{A.179}$$

where the approximated displacement vector is

$$\mathbf{s} = (\mathbf{r}_i(t) + h\mathbf{q}_i(t)) - (\mathbf{r}_j(t) + h\mathbf{q}_j(t)). \tag{A.180}$$

This displacement's vector should obey the given constraints. Evaluating this gives two options to follow with respect to the given tolerance constant ε. If the following relation,

$$|\mathbf{s}|^2 - d_{ij}^2 \leqslant \varepsilon, \tag{A.181}$$

holds, the calculation's loop should be restarted with the newly chosen constraint to evaluate. If the relation does not hold, the position vector should be corrected. To be

more specific, the calculation of G parameter can be done by using the following equation

$$G = \frac{(s^2 - d_{ij}^2)}{2h} \left[\mathbf{s} \cdot \mathbf{r}_{ij}(t)\right] \left(\frac{1}{m_i} + \frac{1}{m_j}\right) . \tag{A.182}$$

This G parameter is used as a correction variable in order to satisfy the range of the given tolerance. It derived from the following proposed equations

$$\mathbf{r}_i^T = \mathbf{r}_i(t) + h \left[\mathbf{q}_i - G\frac{\mathbf{r}_{ij}(t)}{m_i}\right] \tag{A.183}$$

and

$$\mathbf{r}_j^T = \mathbf{r}_j(t) + h \left[\mathbf{q}_j + G\frac{\mathbf{r}_{ij}(t)}{m_j}\right] \tag{A.184}$$

in which describing position vector of particle i and j at time $t + h$ respectively. In order to satisfy the following relation

$$\left|\mathbf{r}_i^T - \mathbf{r}_j^T\right|^2 = d_{ij}^2. \tag{A.185}$$

Solving Eq.[A.183] and Eq.[A.184] by concerning the relation written in Eq.[A.181] gives Eq.[A.182]; here, the second order of G is omitted. The resulted G is used to correct $\mathbf{q}_i$ as can be written in the following

$$\begin{aligned} \mathbf{q}_i &= \mathbf{q}_i - G\frac{\mathbf{r}_{ij}(t)}{m_i} \\ \mathbf{q}_j &= \mathbf{q}_j - G\frac{\mathbf{r}_{ij}(t)}{m_j} . \end{aligned} \tag{A.186}$$

Position's vector is then recalculated by using Eq.[A.184] and Eq.[A.185]. After this step, the calculation loop is restarted all over again with the new chosen constraint, and repeated until all of the given constraints are corrected.

So far, the corrections were done in the position vector. To start with the velocity, let

$$\dot{\mathbf{r}}_i(t + h) = \mathbf{q}_i + \frac{h}{2m_i}\mathbf{F}_i(t + h) \tag{A.187}$$

be the equation to be used in the calculation after a contraint to be evaluated is depicted. Subsequently, the dot product of position's vector and its time derivative (Eq.[A.187]) at time $t + h$ is employed to satisfy

$$\dot{\mathbf{r}}_{ij}(t + h) \cdot \mathbf{r}_{ij}(t + h) \leqslant \varepsilon . \tag{A.188}$$

If the relation written in Eq.[A.188] holds, the chosen constraint is no longer need a correction. The calculation process is then go backward to the picking stage of the next constraint's variable to evaluate. Subsequently, if the condition is not satisfied in this

newly chosen constraint's variable, the corrections would be done by performing the following equation,

$$\gamma = \mathbf{r}_{ij}(t+h) \cdot \frac{[\dot{\mathbf{r}}_i(t+h) - \dot{\mathbf{r}}_j(t+h)]}{d_{ij}^2} \left(\frac{1}{m_i} + \frac{1}{m_j} \right) , \tag{A.189}$$

at first. Next, the replacement of $\dot{\mathbf{r}}_i(t+h)$ by

$$\dot{\mathbf{r}}_i^T = \dot{\mathbf{r}}_i(t+h) - \frac{\gamma}{m_i} \mathbf{r}_{ij}(t+h) , \tag{A.190}$$

and $\dot{\mathbf{r}}_j(t+h)$ by

$$\dot{\mathbf{r}}_j^T = \dot{\mathbf{r}}_j(t+h) + \frac{\gamma}{m_j} \mathbf{r}_{ij}(t+h) . \tag{A.191}$$

should be done. Next stage would be to evaluate other velocity constraint and restarting the loop. Termination will be employed after all of the constraints are evaluated.

B

Diffusion Coefficient Code

The following code is an example code to calculate mean square displacement of one vacancy hydrate C_1 system.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define NATOM          53   // number of guest molecules
#define NSTAT      500001   // simulation period
#define NTCF        20000   // correlation time

double **ecalloc_dmat(int nrow, int ncol);

int min(int a, int b) {
  if (a < b) {
    return a;
  } else {
    return b;
  }
}

double *pbc(double ax,double ay,double az,
            double bx,double by,double bz,
            double cx,double cy,double cz,
            double x,double y,double z)
{
  int i;
  double  Minv[9];
  double  detM, boxa, boxb, boxc;
  static  double dr[3];
  detM    = ax*by*cz - ax*cy*bz
            + ay*bz*cx -ay*bx*cz
            + az*bx*cy - az*by*cx ;
  Minv[0] = by*cz - cy*bz ; Minv[1] = bz*cx - bx*cz ;
  Minv[2] = bx*cy - by*cx ; Minv[3] = cy*az - cz*ay ;
  Minv[4] = cz*ax - cx*az ; Minv[5] = cx*ay - ax*cy ;
  Minv[6] = ay*bz - az*by ; Minv[7] = az*bx - ax*bz ;
  Minv[8] = ax*by - bx*ay ;
  for(i=0;i<9;i++){
```

```
    Minv[i]/= detM;
  }
  boxa = Minv[0]*x + Minv[1]*y + Minv[2]*z ;
  boxb = Minv[3]*x + Minv[4]*y + Minv[5]*z ;
  boxc = Minv[6]*x + Minv[7]*y + Minv[8]*z ;
  boxa -= rint(boxa);
  boxb -= rint(boxb);
  boxc -= rint(boxc);
  dr[0] = ax*boxa + bx*boxb +cx*boxc;
  dr[1] = ay*boxa + by*boxb +cy*boxc;
  dr[2] = az*boxa + bz*boxb +cz*boxc;
return dr;
}

int main(void)
{
  int i, j,k, istat, itcf, isample;
  int nstat=NSTAT, natom=NATOM;
  int ntcf=NTCF,nsample;
  char an[15],bn[15],cn[15];
  double cf[NTCF+1],norm[NTCF+1];
  int iatom,katom;
  double **xg,**yg,**zg;
  double *xlo,*ylo,*zlo;
  double *xhi,*yhi,*zhi;
  double *xy,*xz,*yz;
  double rx[natom],ry[natom],rz[natom];
  double *drr;
  double dx,dy,dz,time;

  FILE *Fco, *Fbx, *Fot;

  Fco  = fopen("FII_eguest.xyz", "r"); // guest trajectory data
  Fbx  = fopen("boxx.dat", "r");       // box's dimension evolution
  Fot  = fopen("Res_MSD.dat", "w+");    // results

  xg   = ecalloc_dmat(nstat, natom);
  yg   = ecalloc_dmat(nstat, natom);
  zg   = ecalloc_dmat(nstat, natom);
  xlo  = (double*)calloc(nstat,sizeof(double));
  xhi  = (double*)calloc(nstat,sizeof(double));
  ylo  = (double*)calloc(nstat,sizeof(double));
  yhi  = (double*)calloc(nstat,sizeof(double));
  zlo  = (double*)calloc(nstat,sizeof(double));
  zhi  = (double*)calloc(nstat,sizeof(double));
  xy   = (double*)calloc(nstat,sizeof(double));
  xz   = (double*)calloc(nstat,sizeof(double));
  yz   = (double*)calloc(nstat,sizeof(double));

  /*  Read coordinate file  */
  for (i = 0; i < nstat; i++) {
    fscanf(Fco,"%d",&iatom);
    fscanf(Fco,"%s %s %s",an,bn,cn);
```

```
    for (j = 0; j < natom; j++){
      fscanf(Fco,"%d %lf %lf %lf",
             &katom, &xg[i][j], &yg[i][j], &zg[i][j]);
    }
  }
  /*  Read box geometry file  */
  for (i=0;i<nstat;i++){
    fscanf(Fbx,"%lf %lf %lf", &xlo[i],&xhi[i],&xy[i]);
    fscanf(Fbx,"%lf %lf %lf", &ylo[i],&yhi[i],&xz[i]);
    fscanf(Fbx,"%lf %lf %lf", &zlo[i],&zhi[i],&yz[i]);
  }
  for (i = 0; i <= ntcf; i++) {
    cf[i] = 0.0; norm[i] = 0.0;
  }
  /*  k gives time origin  */
  for (k = 0; k < nstat; k++) {
    for (i = 0; i < natom; i++) {
      rx[i] = 0.0; ry[i] = 0.0; rz[i] = 0.0;
    }
    nsample = min(nstat, k+ntcf+1);
    for (isample = k+1; isample < nsample; isample++) {
      itcf = isample - k;
      for (i = 0; i < natom; i++) {
        dx = xg[isample][i] - xg[isample-1][i];
        dy = yg[isample][i] - yg[isample-1][i];
        dz = zg[isample][i] - zg[isample-1][i];
        drr = pbc(xhi[k]-xlo[k],0,0,
                  xy[k],yhi[k]-ylo[k],0,
                  xz[k],yz[k],zhi[k]-zlo[k],dx,dy,dz);
        rx[i] += drr[0] ; ry[i] += drr[1] ; rz[i] += drr[2];
        cf[itcf] += rx[i]*rx[i] + ry[i]*ry[i] + rz[i]*rz[i];
      }
      norm[itcf] += 1.0;
    }
  }
  /*  Normalize correlation function  */
  for (i = 1; i <= ntcf; i++) cf[i] /= (norm[i]*natom);
  for (i = 0; i <= ntcf; i++) {
    time = (i + 1.0)*100;                       // in femtosecond
    fprintf(Fot, "%6.2f  %f\n", time, cf[i]);
  }
  fclose(Fco);fclose(Fbx);fclose(Fot);
  free(xlo); free(xhi);
  free(ylo); free(yhi);
  free(zlo); free(zhi);
  free(xy); free(xz); free(yz);
  return 0;
}

double **ecalloc_dmat(int nrow, int ncol)
{
  int i;
  double **p;
```

```
  /*  allocate pointers to rows  */
  p = (double **) malloc(nrow*sizeof(double *));
  if (!p) {
    printf("Error: can not allocate "
           "memory for double matrix (%lu bytes)\n",
           nrow*sizeof(double *));
    exit (2);
  }
  /*  allocate rows  */
  p[0] = (double *) calloc(nrow*ncol, sizeof(double));
  if (!p[0]) {
    printf("Error: can not allocate "
           "memory for arrays for "
           "double matrix (%lu bytes)\n",
           nrow*ncol*sizeof(double));
    exit (2);
  }
  /*  set pointers to rows  */
  for (i = 1; i < nrow; ++i) p[i] = p[i-1] + ncol;
  return p;
}
```

The following is a log file from linear curve fitting performed by gnuplot software. The data is taken from previous MSD code's output. The diffusion coefficient is then calculated from a variable written in "Final set parameters", divided by 6 and converted to unit cm^2/s.

```
*******************************************************************************
Thu Jan 11 10:08:59 2018


FIT:    data read from 'Res_MSD.dat' u 1:2
        format = x:z
        #datapoints = 20001
        residuals are weighted equally (unit weight)

function used for fitting: f(x)
        f(x) = a*x +b
fitted parameters initialized with current variable values

iter      chisq       delta/lim  lambda   a             b
   0 2.6672681631e+16   0.00e+00  8.17e+05    1.000000e+00   1.000000e+00
   8 4.0820021070e+00  -5.19e-05  8.17e-03    2.180807e-07   3.459812e-01

After 8 iterations the fit converged.
final sum of squares of residuals : 4.082
rel. change during last iteration : -5.19026e-10

degrees of freedom    (FIT_NDF)                        : 19999
rms of residuals      (FIT_STDFIT) = sqrt(WSSR/ndf)    : 0.0142867
variance of residuals (reduced chisquare) = WSSR/ndf   : 0.00020411

Final set of parameters            Asymptotic Standard Error
```

```
======================            ==========================
a                = 2.18081e-07      +/- 1.75e-10     (0.08023%)
b                = 0.345981         +/- 0.000202     (0.0584%)

correlation matrix of the fit parameters:
                 a      b
a                1.000
b               -0.866  1.000
```

C

Potential Energy Map of Trial Guest Position Calculation

In the case of hydrate C_1 system, the following codes that are arranged sequentially generate potential energy map on basal plane. Firstly, the code shown below give grid points on the chosen plane.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#define NATOM      53
double get_dist(double xa,double ya,double za){
  return sqrt(pow(xa,2)+pow(ya,2)+pow(za,2));
}
int main(void)
{
  int i,j;
  int natom=NATOM;
  char an[15],bn[15],cn[15];
  int iatom,katom;
  int n1grid,n2grid,adtn;
  double xg[natom],yg[natom],zg[natom];
  double x,y,z;
  double dx1,dy1,dz1,r_d1;
  double dx2,dy2,dz2,r_d2;
  double dx3,dy3,dz3,r_d3;
  double ux1,uy1,uz1;
  double ux2,uy2,uz2;
  double ux3,uy3,uz3;
  FILE *Fco,*Fot,*Fot1,*Fot2;
  n1grid = 300;
  n2grid = 300;
  adtn   = 200;
  Fco  = fopen("FII_36619.xyz", "r");
  /* grid on optical plane */
  Fot  = fopen("FII_grid_opt.dat", "w+");
  /* grid on basal plane   */
  Fot1 = fopen("FII_grid_bas.dat", "w+");
  /* grid on other basal plane  */
```

```
  Fot2 = fopen("FII_grid_bas2.dat", "w+");
  fscanf(Fco,"%d",&iatom);
  fscanf(Fco,"%s %s %s",an,bn,cn);
  for(i=0;i<natom;i++){
    fscanf(Fco,"%d %lf %lf %lf",&katom,&xg[i],&yg[i],&zg[i]);
  }
/* optical plane : (1) --> 4 --> 40 --> 37 */
  dx1 = xg[4] - xg[1]; dx2 = xg[37] - xg[1];
  dy1 = yg[4] - yg[1]; dy2 = yg[37] - yg[1];
  dz1 = zg[4] - zg[1]; dz2 = zg[37] - zg[1];
  r_d1 = get_dist(dx1,dy1,dz1);
  r_d2 = get_dist(dx2,dy2,dz2);
  ux1 = dx1/r_d1; ux2 = dx2/r_d2;
  uy1 = dy1/r_d1; uy2 = dy2/r_d2;
  uz1 = dz1/r_d1; uz2 = dz2/r_d2;
  for(i=0;i<adtn;i++){
     for(j=0;j<n2grid;j++){
        x = (xg[1]-(i*0.1*ux1)) + (ux2*j*0.1);
        y = (yg[1]-(i*0.1*uy1)) + (uy2*j*0.1);
        z = (zg[1]-(i*0.1*uz1)) + (uz2*j*0.1);
        fprintf(Fot,"%lf %lf %lf\n",x,y,z);
     }
  }
  for(j=0;j<adtn;j++){
     for(i=0;i<n2grid;i++){
        x = (xg[1]+(i*0.1*ux1)) - (ux2*j*0.1);
        y = (yg[1]+(i*0.1*uy1)) - (uy2*j*0.1);
        z = (zg[1]+(i*0.1*uz1)) - (uz2*j*0.1);
        fprintf(Fot,"%lf %lf %lf\n",x,y,z);
     }
  }
  for(i=0;i<adtn;i++){
     for(j=0;j<adtn;j++){
        x = (xg[1]-(i*0.1*ux1)) - (ux2*j*0.1);
        y = (yg[1]-(i*0.1*uy1)) - (uy2*j*0.1);
        z = (zg[1]-(i*0.1*uz1)) - (uz2*j*0.1);
        fprintf(Fot,"%lf %lf %lf\n",x,y,z);
     }
  }
  for(i=0;i<n1grid;i++){
     for(j=0;j<n2grid;j++){
        x = (xg[1]+(i*0.1*ux1)) + (ux2*j*0.1);
        y = (yg[1]+(i*0.1*uy1)) + (uy2*j*0.1);
        z = (zg[1]+(i*0.1*uz1)) + (uz2*j*0.1);
        fprintf(Fot,"%lf %lf %lf\n",x,y,z);
     }
  }
 /* basal plane 1 : (1) --> 4 --> 16 --> 13 */
  dx2 = xg[13] - xg[1];
  dy2 = yg[13] - yg[1];
  dz2 = zg[13] - zg[1];
  r_d2 = get_dist(dx2,dy2,dz2);
  ux2 = dx2/r_d2;
```

```
uy2 = dy2/r_d2;
uz2 = dz2/r_d2;
for(i=0;i<adtn;i++){
   for(j=0;j<n2grid;j++){
      x = (xg[1]-(i*0.1*ux1)) + (ux2*j*0.1);
      y = (yg[1]-(i*0.1*uy1)) + (uy2*j*0.1);
      z = (zg[1]-(i*0.1*uz1)) + (uz2*j*0.1);
      fprintf(Fot1,"%lf %lf %lf\n",x,y,z);
   }
}
for(j=0;j<adtn;j++){
   for(i=0;i<n2grid;i++){
      x = (xg[1]+(i*0.1*ux1)) - (ux2*j*0.1);
      y = (yg[1]+(i*0.1*uy1)) - (uy2*j*0.1);
      z = (zg[1]+(i*0.1*uz1)) - (uz2*j*0.1);
      fprintf(Fot1,"%lf %lf %lf\n",x,y,z);
   }
}
for(i=0;i<adtn;i++){
   for(j=0;j<adtn;j++){
      x = (xg[1]-(i*0.1*ux1)) - (ux2*j*0.1);
      y = (yg[1]-(i*0.1*uy1)) - (uy2*j*0.1);
      z = (zg[1]-(i*0.1*uz1)) - (uz2*j*0.1);
      fprintf(Fot1,"%lf %lf %lf\n",x,y,z);
   }
}
for(i=0;i<n1grid;i++){
   for(j=0;j<n2grid;j++){
      x = (xg[1]+(i*0.1*ux1)) + (ux2*j*0.1);
      y = (yg[1]+(i*0.1*uy1)) + (uy2*j*0.1);
      z = (zg[1]+(i*0.1*uz1)) + (uz2*j*0.1);
      fprintf(Fot1,"%lf %lf %lf\n",x,y,z);
   }
}
/* basal plane (2): 1 --> 37 --> 49 --> 13 */
dx1 = xg[37] - xg[1]; dx2 = xg[13] - xg[1];
dy1 = yg[37] - yg[1]; dy2 = yg[13] - yg[1];
dz1 = zg[37] - zg[1]; dz2 = zg[13] - zg[1];
r_d1 = get_dist(dx1,dy1,dz1);
r_d2 = get_dist(dx2,dy2,dz2);
ux1 = dx1/r_d1; ux2 = dx2/r_d2;
uy1 = dy1/r_d1; uy2 = dy2/r_d2;
uz1 = dz1/r_d1; uz2 = dz2/r_d2;
for(i=0;i<adtn;i++){
   for(j=0;j<n2grid;j++){
      x = (xg[1]-(i*0.1*ux1)) + (ux2*j*0.1);
      y = (yg[1]-(i*0.1*uy1)) + (uy2*j*0.1);
      z = (zg[1]-(i*0.1*uz1)) + (uz2*j*0.1);
      fprintf(Fot2,"%lf %lf %lf\n",x,y,z);
   }
}
for(j=0;j<adtn;j++){
   for(i=0;i<n2grid;i++){
```

```
        x = (xg[1]+(i*0.1*ux1)) - (ux2*j*0.1);
        y = (yg[1]+(i*0.1*uy1)) - (uy2*j*0.1);
        z = (zg[1]+(i*0.1*uz1)) - (uz2*j*0.1);
        fprintf(Fot2,"%lf %lf %lf\n",x,y,z);
     }
  }
  for(i=0;i<adtn;i++){
     for(j=0;j<adtn;j++){
        x = (xg[1]-(i*0.1*ux1)) - (ux2*j*0.1);
        y = (yg[1]-(i*0.1*uy1)) - (uy2*j*0.1);
        z = (zg[1]-(i*0.1*uz1)) - (uz2*j*0.1);
        fprintf(Fot2,"%lf %lf %lf\n",x,y,z);
     }
  }
  for(i=0;i<n1grid;i++){
     for(j=0;j<n2grid;j++){
        x = (xg[1]+(i*0.1*ux1)) + (ux2*j*0.1);
        y = (yg[1]+(i*0.1*uy1)) + (uy2*j*0.1);
        z = (zg[1]+(i*0.1*uz1)) + (uz2*j*0.1);
        fprintf(Fot2,"%lf %lf %lf\n",x,y,z);
     }
  }
fclose(Fco);fclose(Fot);fclose(Fot1); fclose(Fot2);
return 0;
}
```

Secondly, the following code move the tagged guest hydrogen molecule on each grid point and subsequently calculate total potential energy from these trial position. The grid point is obtained from the first code.

```
/* -------------------------------------------------
this program calculates Potential Energy
with respect to the movement of Guest Atom No.1
-------------------------------------------------*/
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define NATOM      1025  /* number of atoms in system */
#define NWAT        324  /* number of water molecules */
#define NGRID       500  /* number of grid */
#define NGUEST       53  /* number of guest molecules */

double **ecalloc_dmat(int nrow, int ncol);

double get_Elj(double e,double s,double r){
  double rinv = 1.0/r;
  return (4.0*e*(pow(rinv*s,12.0) - pow(rinv*s,6.0)));
}

double get_dist(double xa,double ya,double za){
```

```
  return sqrt(pow(xa,2)+pow(ya,2)+pow(za,2));
}

double *pbc(double ax,double ay,double az,
            double bx,double by,double bz,
            double cx,double cy,double cz,
            double x,double y,double z)
{
  int i;
  double  Minv[9];
  double  detM, boxa, boxb, boxc;
  static  double dr[3];

  detM    = ax*by*cz - ax*cy*bz
          + ay*bz*cx -ay*bx*cz
          + az*bx*cy - az*by*cx ;
  Minv[0] = by*cz - cy*bz ; Minv[1] = bz*cx - bx*cz ;
  Minv[2] = bx*cy - by*cx ; Minv[3] = cy*az - cz*ay ;
  Minv[4] = cz*ax - cx*az ; Minv[5] = cx*ay - ax*cy ;
  Minv[6] = ay*bz - az*by ; Minv[7] = az*bx - ax*bz ;
  Minv[8] = ax*by - bx*ay ;
  for(i=0;i<9;i++){
    Minv[i]/= detM;
  }
  boxa = Minv[0]*x + Minv[1]*y + Minv[2]*z ;
  boxb = Minv[3]*x + Minv[4]*y + Minv[5]*z ;
  boxc = Minv[6]*x + Minv[7]*y + Minv[8]*z ;
  boxa -= rint(boxa);
  boxb -= rint(boxb);
  boxc -= rint(boxc);
  dr[0] = ax*boxa + bx*boxb +cx*boxc;
  dr[1] = ay*boxa + by*boxb +cy*boxc;
  dr[2] = az*boxa + bz*boxb +cz*boxc;
return dr;
}

int main(void)
{
  int i,j;
  int nwat=NWAT,natom=NATOM;
  int ngrid=NGRID,nguest=NGUEST;
  double sigm_OO,sigm_HH;
  double eps_OO,eps_HH;
  double sigm_OH,eps_OH;
  char an[15],bn[15],cn[15];
  int iatom,katom;
  double xg[natom],yg[natom],zg[natom];
  double xc[ngrid*ngrid],yc[ngrid*ngrid],zc[ngrid*ngrid];
  double Ep[ngrid*ngrid];
  double dx,dy,dz,r_d;
  double xlo,xhi,ylo,yhi,zlo,zhi;
  double xy,xz,yz;
  double *dr;
```

```
int nwatmol;

/* box dimension */
xlo = -0.042492;xhi= 23.382492;xy = -9.193353;
ylo = -0.039088;yhi= 21.509088;xz =  5.038279;
zlo = -0.029839;zhi= 16.419839;yz =  7.637709;

FILE *Fco, *Fgrd, *Fot;

sigm_OO = 3.1668;        /* O-O sigm (Angstrom) */
sigm_HH = 2.96;          /* H-H sigm (Angstrom) */
eps_HH  = 0.067962;      /* H-H eps */
eps_OO  = 0.21084;       /* O-O eps */
sigm_OH = (sigm_OO+sigm_HH)/2;
eps_OH  = sqrt(eps_HH*eps_OO);

Fco  = fopen("../FII_36619comp.xyz","r");
Fgrd = fopen("FII_grid_bas2.dat","r");
Fot  = fopen("EP_bas2_profile.dat","w+");

fscanf(Fco,"%d",&iatom);
fscanf(Fco,"%s %s %s",an,bn,cn);
for(i = 0; i < natom; i++){
  fscanf(Fco,"%d %lf %lf %lf",&katom,&xg[i],&yg[i],&zg[i]);
}
for(i=0;i<ngrid*ngrid;i++){
  fscanf(Fgrd,"%lf %lf %lf",&xc[i],&yc[i],&zc[i]);
}
for(j=0;j<ngrid*ngrid;j++){
   Ep[j]=0.0;
}
for(j=0;j<ngrid*ngrid;j++){
   for(i=0;i<nwat;i++){
      dx  = xc[j] - xg[i*3];
      dy  = yc[j] - yg[i*3];
      dz  = zc[j] - zg[i*3];
      dr  = pbc(xhi-xlo,0,0,xy,yhi-ylo,0,
                xz,yz,zhi-zlo,dx,dy,dz);
      r_d = get_dist(dr[0],dr[1],dr[2]);
      Ep[j] += get_Elj(eps_OH,sigm_OH,r_d);
   }
}
nwatmol = 3*nwat;
for(j=0;j<ngrid*ngrid;j++){
   for(i=0;i<nguest;i++){
      if(i!=1){
        dx  = xc[j] - xg[nwatmol+i];
        dy  = yc[j] - yg[nwatmol+i];
        dz  = zc[j] - zg[nwatmol+i];
        dr  = pbc(xhi-xlo,0,0,xy,yhi-ylo,0,
                  xz,yz,zhi-zlo,dx,dy,dz);
        r_d = get_dist(dr[0],dr[1],dr[2]);
        Ep[j] += get_Elj(eps_HH,sigm_HH,r_d);
```

```
        }
      }
  }
  for(j=0;j<ngrid*ngrid;j++){
    fprintf(Fot,"%lf %lf %lf %lf\n",xc[j],yc[j],zc[j],Ep[j]);
  }
  fclose(Fco);fclose(Fgrd);fclose(Fot);
  return 0;
}


double **ecalloc_dmat(int nrow, int ncol)
{
  int i;
  double **p;
  /*  allocate pointers to rows  */
  p = (double **) malloc(nrow*sizeof(double *));
  if (!p) {
    printf("Error: can not allocate memory "
           "for double matrix (%lu bytes)\n",
           nrow*sizeof(double *));
    exit (2);
  }
  /*  allocate rows  */
  p[0] = (double *) calloc(nrow*ncol, sizeof(double));
  if (!p[0]) {
    printf("Error: can not allocate memory "
           "for arrays for double matrix (%lu bytes)\n",
           nrow*ncol*sizeof(double));
    exit (2);
  }
  /*  set pointers to rows  */
  for (i = 1; i < nrow; ++i) p[i] = p[i-1] + ncol;
  return p;
}
```

To finalize the calculation, all of the atoms are projected on to the plane by using the code shown below.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define NATOM      1025  /* number of atoms in system */
#define NWAT        324  /* number of water molecules */
#define NGRID       500  /* number of grid */
#define NGUEST       53  /* number of guest molecules */

double get_dist(double xa,double ya,double za){
  return sqrt(pow(xa,2)+pow(ya,2)+pow(za,2));
}

double *get_cross(double xa, double ya, double za,
                  double xb, double yb, double zb){
```

```
  static double u[3];
  double a,b,c,d;
  a = (ya*zb)-(yb*za);
  b = (xb*za)-(xa*zb);
  c = (xa*yb)-(xb*ya);
  d = sqrt((a*a)+(b*b)+(c*c));
  u[0] = a/d;
  u[1] = b/d;
  u[2] = c/d;
  return u;
}
double get_dot(double x1,double y1,double z1,
               double x2,double y2,double z2){
 return (x1*x2)+(y1*y2)+(z1*z2);
}

int main(void)
{
  int i,j;
  int natom=NATOM;
  int nguest=NGUEST,nwat=NWAT;
  int ngrid=NGRID;
  char an[15],bn[15],cn[15];
  int iatom,katom;
  double xg[natom],yg[natom],zg[natom];
  double xf[ngrid*ngrid],yf[ngrid*ngrid],zf[ngrid*ngrid];
  double Ef[ngrid*ngrid];
  double x[natom],y[natom],z[natom];
  double dx1,dy1,dz1,r_d1;
  double dx2,dy2,dz2,r_d2;
  double ux1,uy1,uz1;
  double ux2,uy2,uz2;
  double *V_nrm2,*R_v;
  double C2,Rx,Ry,Rz,s,C;
  double a,b,t, R[9];
  double Mx,My,Mz;
  double cos_theta,d_rot;
  double xa,ya,za;
  int nwatmol;
  FILE *Fco;
  FILE *FO2,*FH2,*FG2;
  FILE *Febas,*Fbas;

  Fco  = fopen("../FII_36619comp.xyz", "r");
  /* --- I/O for all atom in the system at 3.661.900 fs ---*/
  fscanf(Fco,"%d",&iatom);
  fscanf(Fco,"%s %s %s",an,bn,cn);
  for(i=0;i<natom;i++){
    fscanf(Fco,"%d %lf %lf %lf",&katom,&xg[i],&yg[i],&zg[i]);
  }
  printf("%lf\n",xg[100]);
  Febas= fopen("EP_bas2_profile.dat","r");
  FO2  = fopen("FII_proj_bas2_oxy.dat", "w+");
```

```
FH2  = fopen("FII_proj_bas2_Hyd.dat", "w+");
FG2  = fopen("FII_proj_bas2_Guest.dat", "w+");
Fbas = fopen("Ep_bas2_xy.dat", "w+");

/*Potential Energy on basal plane*/
for(i=0;i<ngrid*ngrid;i++){
  fscanf(Febas,"%lf %lf %lf %lf",&xf[i],&yf[i],&zf[i],&Ef[i]);
}
/* for basal plane ======================================*/
/* basal plane    :  1 --> 4 --> 16 --> 13   */
/* basal plane(2) : 1 --> 37 --> 49 --> 13   */

nwatmol = 3*nwat;
dx1 = xg[nwatmol+37] - xg[nwatmol+1];
dy1 = yg[nwatmol+37] - yg[nwatmol+1];
dz1 = zg[nwatmol+37] - zg[nwatmol+1];
dx2 = xg[nwatmol+13] - xg[nwatmol+1];
dy2 = yg[nwatmol+13] - yg[nwatmol+1];
dz2 = zg[nwatmol+13] - zg[nwatmol+1];
r_d1 = get_dist(dx1,dy1,dz1);
r_d2 = get_dist(dx2,dy2,dz2);
ux1 = dx1/r_d1; ux2 = dx2/r_d2;
uy1 = dy1/r_d1; uy2 = dy2/r_d2;
uz1 = dz1/r_d1; uz2 = dz2/r_d2;
V_nrm2 = get_cross(ux1,uy1,uz1,ux2,uy2,uz2);
Mx = V_nrm2[0];
My = V_nrm2[1];
Mz = V_nrm2[2];
printf("2 %lf %lf %lf\n",Mx,My,Mz);
C2 = (Mx*xg[nwatmol+4])
     +(My*yg[nwatmol+4])
     +(Mz*zg[nwatmol+4]);
/*=========================================================*/
/* 0. find unit vector of rotation axis */
cos_theta = get_dot(Mx,My,Mz,0,0,1);
R_v = get_cross(Mx,My,Mz,0,0,1);
d_rot = get_dist(R_v[0],R_v[1],R_v[2]);
Rx = R_v[0]/d_rot;
Ry = R_v[1]/d_rot;
Rz = R_v[2]/d_rot;
s  = sqrt(1-(cos_theta*cos_theta));
C  = 1-cos_theta;
 /* 1. Find rotation matrix---------------------------------*/
R[0] = (Rx*Rx*C) + cos_theta;  R[1] = (Rx*Ry*C)-(Rz*s)     ;
R[2] = (Rx*Rz*C)+(Ry*s);
R[3] = (Ry*Rx*C)+(Rz*s)     ;  R[4] = (Ry*Ry*C)+ cos_theta ;
R[5] = (Ry*Rz*C)-(Rx*s);
R[6] = (Rz*Rx*C)-(Ry*s)     ;  R[7] = (Rz*Ry*C)+(Rx*s)     ;
R[8] = (Rz*Rz*C)+ cos_theta;
/* ------------------------------------------------------ */
/* 2. project all atoms to basal plane */
a = (Mx*Mx)+(My*My)+(Mz*Mz);
for(i=0;i<natom;i++){
```

```
    b = C2 - ((Mx*xg[i])+(My*yg[i])+(Mz*zg[i]));
    t = b/a;
    x[i] = (Mx*t)+xg[i];
    y[i] = (My*t)+yg[i];
    z[i] = (Mz*t)+zg[i];
  }
  /* print the projected coordinates */
  /* Oxygen water */
  for(i=0;i<nwat;i++){
    xa = R[0]*x[i*3] + R[1]*y[i*3] + R[2]*z[i*3];
    ya = R[3]*x[i*3] + R[4]*y[i*3] + R[5]*z[i*3];
    za = R[6]*x[i*3] + R[7]*y[i*3] + R[8]*z[i*3];
    fprintf(FO2,"%lf %lf %lf\n",xa,ya,za);
  }
  /* Hydrogen water */
  for(i=0;i<nwat;i++){
    for(j=1;j<3;j++){
      xa = R[0]*x[i*3+j] + R[1]*y[i*3+j] + R[2]*z[i*3+j];
      ya = R[3]*x[i*3+j] + R[4]*y[i*3+j] + R[5]*z[i*3+j];
      za = R[6]*x[i*3+j] + R[7]*y[i*3+j] + R[8]*z[i*3+j];
      fprintf(FH2,"%lf %lf %lf\n",xa,ya,za);
    }
  }
  /* Guest */;
  for(i=0;i<nguest;i++){
    xa = R[0]*x[nwatmol+i] + R[1]*y[nwatmol+i] + R[2]*z[nwatmol+i];
    ya = R[3]*x[nwatmol+i] + R[4]*y[nwatmol+i] + R[5]*z[nwatmol+i];
    za = R[6]*x[nwatmol+i] + R[7]*y[nwatmol+i] + R[8]*z[nwatmol+i];
    fprintf(FG2,"%lf %lf %lf\n",xa,ya,za);
  }
  /* Potential Energy */
  for(i=0;i<ngrid*ngrid;i++){
    xa = R[0]*xf[i] + R[1]*yf[i] + R[2]*zf[i];
    ya = R[3]*xf[i] + R[4]*yf[i] + R[5]*zf[i];
    za = R[6]*xf[i] + R[7]*yf[i] + R[8]*zf[i];
    fprintf(Fbas,"%lf %lf %lf %lf\n",xa,ya,za,Ef[i]);
  }
  fclose(Fco);
  fclose(FO2);fclose(FH2);fclose(FG2);
  fclose(Febas);fclose(Fbas);
  return 0;
}
```

D

Infra Red Spectra Calculation Code

The following code [108] calculates IR spectra from LAMMPS output file of system's total dipole moment.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import fftpack
from scipy import signal

# Inputs
# 1 to calculate it, 2 to load a pre-calculated one
autocorrelation_option = 1
T = 291. # K
zoom_wavenum = 4000. # cm^-1
# ----------------------------------
# since you only want to fft the part
# that has meaningful statistics,
# ----------------------------------
fraction_autocorrelation_function_to_fft = 0.01
# ------------------------------------------------
# and at 1% of the trajectory all datapoints have
# at least 100 non-ish-correlated counts
# ------------------------------------------------
input_dipole_file = 'FII_50nsdip.txt'
output_autocorrelation_file = 'FII_autocorr.txt'
# ------------------------------------------------
# Constants
boltz = 1.38064852E-23 # m^2 kg s^-2 K^-1
lightspeed = 299792458. # m s^-1
reduced_planck = 1.05457180013E-34 # kg m^2 s^-1
# ------------------------------------------------
# Get autocorrelation function
# ------------------------------------------------
# Calculate autocorrelation function
if autocorrelation_option == 1:
    # Load data
    time, dipole_x, dipole_y, dipole_z
       = np.loadtxt(input_dipole_file, skiprows=2,
         usecols=(1,2,3,4), unpack=True)
    # Do calculation
    # Note that this method of calculating
```

```
    # an autocorrelation function is very fast,
    # but it can be difficult to follow.
    print("Calculating autocorrelation function.")
    # Shift the array
    if len(time) % 2 == 0:
        dipole_x_shifted = np.zeros(len(time)*2)
        dipole_y_shifted = np.zeros(len(time)*2)
        dipole_z_shifted = np.zeros(len(time)*2)
    else:
        dipole_x_shifted = np.zeros(len(time)*2-1)
        dipole_y_shifted = np.zeros(len(time)*2-1)
        dipole_z_shifted = np.zeros(len(time)*2-1)
    dipole_x_shifted[len(time)//2:len(time)//2+len(time)] = dipole_x
    dipole_y_shifted[len(time)//2:len(time)//2+len(time)] = dipole_y
    dipole_z_shifted[len(time)//2:len(time)//2+len(time)] = dipole_z
    # Convolute the shifted array with the flipped array,
    # which is equivalent to performing a correlation
    autocorr_x_full = (signal.fftconvolve(dipole_x_shifted,
                      dipole_x[::-1], mode='same')[(-len(time)):]
                      / np.arange(len(time), 0, -1))
    autocorr_y_full = (signal.fftconvolve(dipole_y_shifted,
                      dipole_y[::-1], mode='same')[(-len(time)):]
                      / np.arange(len(time), 0, -1))
    autocorr_z_full = (signal.fftconvolve(dipole_z_shifted,
                      dipole_z[::-1], mode='same')[(-len(time)):]
                      / np.arange(len(time), 0, -1))
    autocorr_full = autocorr_x_full + autocorr_y_full + autocorr_z_full
    # Truncate the autocorrelation array
    autocorr = autocorr_full[:int(len(time)
               * fraction_autocorrelation_function_to_fft)]
    print("Finished with autocorrelation function calculation.")
    np.savetxt(output_autocorrelation_file,
               np.column_stack((time[:len(autocorr)], autocorr)),
               header='Time(fs) Autocorrelation(e*Ang)')
# Load pre-calculated autocorrelation function
elif autocorrelation_option == 2:
    time, autocorr = np.loadtxt(output_autocorrelation_file
                     , skiprows=1, unpack=True)
else:
    print("Not a valid option for 'autocorrelation_option'.")

# converts time from femtoseconds to seconds
timestep = (time[1]-time[0]) * 1.E-15
# ------------------------------------------------
# Calculate spectra
# Note that intensities are relative, and so can be
# multiplied by a constant to compare to experiment.
# ------------------------------------------------
# Calculate the FFTs of autocorrelation functions
lineshape = fftpack.dct(autocorr, type=1)[1:]
lineshape_frequencies =
     np.linspace(0, 0.5/timestep, len(autocorr))[1:]
# converts to wavenumbers (cm^-1)
```

```
lineshape_frequencies_wn =
     lineshape_frequencies / (100.*lightspeed)

# Calculate spectra
field_description =
     lineshape_frequencies
     * (1. - np.exp(-reduced_planck
     *lineshape_frequencies/(boltz*T)))
# quantum correction per doi.org/10.1021/jp034788u.
# Other options are possible, see doi.org/10.1063/1.441739
# and doi.org/10.1080/00268978500102801.
quantum_correction =
     lineshape_frequencies /
     (1. - np.exp(-reduced_planck
     *lineshape_frequencies/(boltz*T)))
spectra = lineshape * field_description
spectra_qm = spectra * quantum_correction
# Save data
np.savetxt('FII_IRdata.txt',
     np.column_stack((lineshape_frequencies_wn,
     lineshape, field_description,
     quantum_correction, spectra, spectra_qm)),
     header='Frequency(cm^-1), Lineshape, '
     'Field_description, Quantum_correction, '
     'Spectra, Spectra_qm')
# ------------------------------------------------------
# Plots
# ------------------------------------------------------
mask = (lineshape_frequencies_wn >= 0)
  & (lineshape_frequencies_wn <= zoom_wavenum)
```

References

[1] J. Ahn, M. Wahlen, B. L. Deck, E. J. Brook, P. A. Mayewski, K. C. Taylor, and J. W. C. White, "A Record of Atmospheric CO_2 during the Last 40,000 Years from the Siple Dome, Antarctica Ice Core: ICE CORE CO $_2$ DURING THE LAST 40 KYR," *Journal of Geophysical Research: Atmospheres*, vol. 109, pp. n/a–n/a, July 2004.

[2] J. Ahn, M. Headly, M. Wahlen, E. J. Brook, P. A. Mayewski, and K. C. Taylor, "CO_2 Diffusion in Polar Ice: Observations from Naturally Formed CO_2 Spikes in the Siple Dome (Antarctica) Ice Core," *Journal of Glaciology*, vol. 54, no. 187, pp. 685–695, 2008.

[3] O. US EPA, "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016," Jan. 2018.

[4] W. L. Mao, C. A. Koh, and E. D. Sloan, "Clathrate Hydrates Under Pressure," *Physics Today*, vol. 60, p. 42, Oct. 2007.

[5] E. D. S. Jr and C. Koh, *Clathrate Hydrates of Natural Gases, Third Edition*. CRC Press, Sept. 2007. Google-Books-ID: T7LC8ldaVR4C.

[6] V. R. Belosludov, T. M. Inerbaev, R. V. Belosludov, J.-i. Kudoh, and Y. Kawazoe, "Absolute Stability Boundaries of Clathrate Hydrates of Cubic Structure II," *Journal of Supramolecular Chemistry*, vol. 2, pp. 377–383, Aug. 2002.

[7] M. Lasich, A. H. Mohammadi, and D. Ramjugernath, "Factors Influencing Clathrate Hydrate Stability in Equilibrium with Liquid Water: Insights from Information-Based Statistical Analysis," *Journal of Molecular Liquids*, vol. 222, pp. 8–13, Oct. 2016.

[8] E. Sloan, J. Bloys, and S. o. P. Engineers (U.S.), *Hydrate Engineering*. Monograph (Society of Petroleum Engineers (U.S.))), Society of Petroleum Engineers, 2000.

[9] K. A. Kvenvolden, "Potential Effects of Gas Hydrate on Human Welfare," *Proceedings of the National Academy of Sciences*, vol. 96, pp. 3420–3426, Mar. 1999.

[10] M. D. Max, ed., *Natural Gas Hydrate: In Oceanic and Permafrost Environments*. Coastal Systems and Continental Margins, Springer Netherlands, 2003.

[11] Y. He, E. S. J. Rudolph, P. L. J. Zitha, and M. Golombok, "Kinetics of CO_2 and Methane Hydrate Formation: An Experimental Analysis in the Bulk phase," *Fuel*, vol. 90, pp. 272–279, Jan. 2011.

[12] V. V. Struzhkin, B. Militzer, W. L. Mao, H.-k. Mao, and R. J. Hemley, "Hydrogen Storage in Molecular Clathrates," *Chemical Reviews*, vol. 107, pp. 4133–4151, Oct. 2007.

[13] W. L. Mao and H.-k. Mao, "Hydrogen Storage in Molecular Compounds," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 708–710, Jan. 2004.

[14] H. Lee, J.-w. Lee, D. Y. Kim, J. Park, Y.-T. Seo, H. Zeng, I. L. Moudrakovski, C. I. Ratcliffe, and J. A. Ripmeester, "Tuning Clathrate Hydrates for Hydrogen Storage," *Nature*, vol. 434, pp. 743–746, Apr. 2005.

[15] Y. H. Hu and E. Ruckenstein, "Clathrate Hydrogen Hydrate-A Promising Material for Hydrogen Storage," *Angewandte Chemie International Edition*, vol. 45, pp. 2011–2013, Mar. 2006.

[16] N. I. Papadimitriou, I. N. Tsimpanogiannis, C. J. Peters, A. T. Papaioannou, and A. K. Stubos, "Hydrogen Storage in sH Hydrates: A Monte Carlo Study," *The Journal of Physical Chemistry B*, vol. 112, pp. 14206–14211, Nov. 2008.

[17] K. Katsumasa, K. Koga, and H. Tanaka, "On the Thermodynamic Stability of Hydrogen Clathrate Hydrates," *The Journal of Chemical Physics*, vol. 127, p. 044509, July 2007.

[18] T. A. Strobel, C. A. Koh, and E. D. Sloan, "Water Cavities of sH Clathrate Hydrate Stabilized by Molecular Hydrogen," *The Journal of Physical Chemistry B*, vol. 112, pp. 1885–1887, Feb. 2008.

[19] T. A. Strobel, C. J. Taylor, K. C. Hester, S. F. Dec, C. A. Koh, K. T. Miller, and E. D. Sloan, "Molecular Hydrogen Storage in Binary THF − H_2 Clathrate Hydrates," *The Journal of Physical Chemistry B*, vol. 110, pp. 17121–17125, Aug. 2006.

[20] R. Anderson, A. Chapoy, and B. Tohidi, "Phase Relations and Binary Clathrate Hydrate Formation in the System H_2 − THF − H_2O," *Langmuir*, vol. 23, pp. 3440–3444, Mar. 2007.

[21] T. Sugahara, J. C. Haag, A. A. Warntjes, P. S. R. Prasad, E. D. Sloan, C. A. Koh, and A. K. Sum, "Large-Cage Occupancies of Hydrogen in Binary Clathrate Hydrates Dependent on Pressures and Guest Concentrations," *The Journal of Physical Chemistry C*, vol. 114, pp. 15218–15222, Sept. 2010.

[22] W. L. Mao, H.-k. Mao, A. F. Goncharov, V. V. Struzhkin, Q. Guo, J. Hu, J. Shu, R. J. Hemley, M. Somayazulu, and Y. Zhao, "Hydrogen Clusters in Clathrate Hydrate," *Science*, vol. 297, pp. 2247–2249, Sept. 2002.

[23] W. L. Vos, L. W. Finger, R. J. Hemley, and H.-k. Mao, "Novel H_2-H_2O Clathrates at High Pressures," *Physical Review Letters*, vol. 71, pp. 3150–3153, Nov. 1993.

[24] H. P. Veluswamy, R. Kumar, and P. Linga, "Hydrogen Storage in Clathrate Hydrates: Current State of the Art and Future Directions," *Applied Energy*, vol. 122, pp. 112–132, June 2014.

[25] W. L. Vos, L. W. Finger, R. J. Hemley, and H.-k. Mao, "Pressure Dependence of Hydrogen Bonding in a Novel H_2O-H_2 Clathrate," *Chemical Physics Letters*, vol. 257, pp. 524–530, Aug. 1996.

[26] V. F. Petrenko and R. W. Whitworth, *Physics of Ice*. Oxford: Oxford University Press, 2006. OCLC: 455956454.

[27] A. D. Fortes, I. G. Wood, J. P. Brodholt, and L. Vočadlo, "Ab Initio Simulation

of the Ice II Structure," *The Journal of Chemical Physics*, vol. 119, pp. 4567–4572, Aug. 2003.

[28] L. Hakim, K. Koga, and H. Tanaka, "Phase Behavior of Different Forms of Ice Filled with Hydrogen Molecules," *Physical Review Letters*, vol. 104, Mar. 2010.

[29] L. Hakim, K. Koga, and H. Tanaka, "Thermodynamic Stability of Hydrogen Hydrates of Ice Ic and II Structures," *Physical Review B*, vol. 82, Oct. 2010.

[30] D. Londono, W. F. Kuhs, and J. L. Finney, "Enclathration of Helium in Ice II: the First Helium Hydrate," *Nature*, vol. 332, pp. 141–142, Mar. 1988.

[31] Y. A. Dyadin, E. G. Larionov, E. Y. Aladko, A. Y. Manakov, F. V. Zhurko, T. V. Mikina, V. Y. Komarov, and E. V. Grachev, "Clathrate Formation in Water-Noble Gas (Hydrogen) Systems at High Pressures," *Journal of Structural Chemistry*, vol. 40, pp. 790–795, Sept. 1999.

[32] M. A. Carignano, "Formation of Stacking Faults during Ice Growth on Hexagonal and Cubic Substrates," *The Journal of Physical Chemistry C*, vol. 111, pp. 501–504, Jan. 2007.

[33] T. Okuchi, M. Takigawa, J. Shu, H.-k. Mao, R. J. Hemley, and T. Yagi, "Fast Molecular Transport in Hydrogen Hydrates by High-Pressure Diamond Anvil Cell NMR," *Physical Review B*, vol. 75, p. 144104, Apr. 2007.

[34] T. Okuchi, "Collision and Diffusion Dynamics of Dense Molecular Hydrogen by Diamond Anvil Cell Nuclear Magnetic Resonance," *The Journal of Physical Chemistry C*, vol. 116, pp. 2179–2182, Jan. 2012.

[35] M. Russina, E. Kemner, and F. Mezei, "Intra-Cage Dynamics of Molecular Hydrogen Confined in Cages of Two Different Dimensions of Clathrate Hydrates," *Scientific Reports*, vol. 6, July 2016.

[36] S. Liang, D. Liang, N. Wu, L. Yi, and G. Hu, "Molecular Mechanisms of Gas Diffusion in CO $_2$ Hydrates," *The Journal of Physical Chemistry C*, vol. 120, pp. 16298–16304, Aug. 2016.

[37] H. Lo, M.-T. Lee, and S.-T. Lin, "Water Vacancy Driven Diffusion in Clathrate Hydrates: Molecular Dynamics Simulation Study," *The Journal of Physical Chemistry C*, vol. 121, pp. 8280–8289, Apr. 2017.

[38] A. Demurov, R. Radhakrishnan, and B. L. Trout, "Computations of Diffusivities in Ice and CO_2 Clathrate Hydrates Via Molecular Dynamics and Monte Carlo Simulations," *The Journal of Chemical Physics*, vol. 116, pp. 702–709, Jan. 2002.

[39] S. Alavi and J. Ripmeester, "Hydrogen-Gas Migration through Clathrate Hydrate Cages," *Angewandte Chemie International Edition*, vol. 46, pp. 6102–6105, Aug. 2007.

[40] K. Momma and F. Izumi, "*VESTA 3* for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data," *Journal of Applied Crystallography*, vol. 44, pp. 1272–1276, Dec. 2011.

[41] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, Feb. 1996.

[42] J. D. Bernal and R. H. Fowler, "A Theory of Water and Ionic Solution, with Partic-

ular Reference to Hydrogen and Hydroxyl Ions," *The Journal of Chemical Physics*, vol. 1, pp. 515–548, Aug. 1933.

[43] M. I. Ryzhkin, A. V. Klyuev, V. V. Sinitsyn, and I. A. Ryzhkin, "Liquid State of a Hydrogen Bond Network in Ice," *JETP Letters*, vol. 104, pp. 248–252, Aug. 2016.

[44] J. J. Shephard, B. Slater, P. Harvey, M. Hart, C. L. Bull, S. T. Bramwell, and C. G. Salzmann, "Doping-Induced Disappearance of Ice II from Water' s Phase Diagram," *Nature Physics*, Apr. 2018.

[45] T. Nakamura, M. Matsumoto, T. Yagasaki, and H. Tanaka, "Thermodynamic Stability of Ice II and Its Hydrogen-Disordered Counterpart: Role of Zero-Point Energy," *The Journal of Physical Chemistry B*, vol. 120, pp. 1843–1848, Mar. 2016.

[46] B. Kamb, "Ice II. a Proton-Ordered form of Ice," *Acta crystallographica*, vol. 17, no. 11, pp. 1437–1449, 1964.

[47] A. D. Fortes, I. G. Wood, M. Alfredsson, L. Vočadlo, and K. S. Knight, "The Incompressibility and Thermal Expansivity of D $_2$ O Ice II Determined by Powder Neutron Diffraction," *Journal of Applied Crystallography*, vol. 38, pp. 612–618, Aug. 2005.

[48] F. W. Starr, C. A. Angell, and H. E. Stanley, "Prediction of Entropy and Dynamic Properties of Water Below the Homogeneous Nucleation Temperature," *Physica A: Statistical Mechanics and its Applications*, vol. 323, pp. 51–66, May 2003.

[49] L. G. Dowell and A. P. Rinfret, "Low-Temperature Forms of Ice as Studied by X-Ray Diffraction," *Nature*, vol. 188, pp. 1144–1148, Dec. 1960.

[50] L. E. Bove, S. Klotz, A. Paciaroni, and F. Sacchetti, "Anomalous Proton Dynamics in Ice at Low Temperatures," *Physical Review Letters*, vol. 103, p. 165901, Oct. 2009.

[51] W. X. Zhang, C. He, J. S. Lian, and Q. Jiang, "Selected Crystallization of Water as a Function of Size," *Chemical Physics Letters*, vol. 421, pp. 251–255, Apr. 2006.

[52] P. Buchanan, A. K. Soper, H. Thompson, R. E. Westacott, J. L. Creek, G. Hobson, and C. A. Koh, "Search for Memory Effects in Methane Hydrate: Structure of Water Before Hydrate Formation and After Hydrate Decomposition," *The Journal of Chemical Physics*, vol. 123, p. 164507, Oct. 2005.

[53] V. A. Vlasov, "Diffusion Model of Gas Hydrate Dissociation into Ice and Gas: Simulation of the Self-Preservation Effect," *International Journal of Heat and Mass Transfer*, vol. 102, pp. 631–636, Nov. 2016.

[54] S. Gao, W. House, and W. G. Chapman, "NMR/MRI Study of Clathrate Hydrate Mechanisms," *The Journal of Physical Chemistry B*, vol. 109, pp. 19090–19093, Oct. 2005.

[55] M. Matsumoto and H. Tanaka, "On the Structure Selectivity of Clathrate Hydrates," *The Journal of Physical Chemistry B*, vol. 115, pp. 8257–8265, June 2011.

[56] B. Tohidi, K. K. Østergaard, A. Danesh, A. C. Todd, and R. W. Burgass, "Structure-H Gas Hydrates in Petroleum Reservoir Fluids," *The Canadian Journal of Chemical Engineering*, vol. 79, pp. 384–391, June 2001.

[57] M. Dutour Sikirić and M. Deza, "Space Fullerenes: Computer Search for New Frank-

Kasper Structures II," *Structural Chemistry*, vol. 23, pp. 1103–1114, Aug. 2012.

[58] J. S.Loveday and R. J.Nelmes, "High-Pressure Gas Hydrates," *Physical Chemistry Chemical Physics*, vol. 10, no. 7, pp. 937–950, 2008.

[59] K. A. Udachin, G. D. Enright, C. I. Ratcliffe, and J. A. Ripmeester, "Structure, Stoichiometry, and Morphology of Bromine Hydrate," *Journal of the American Chemical Society*, vol. 119, pp. 11481–11486, Nov. 1997.

[60] Y. Huang, C. Zhu, L. Wang, X. Cao, Y. Su, X. Jiang, S. Meng, J. Zhao, and X. C. Zeng, "A New Phase Diagram of Water Under Negative Pressure: The Rise of the Lowest-Density Clathrate s-III," *Science Advances*, vol. 2, pp. e1501010–e1501010, Feb. 2016.

[61] D. M. Amos, M.-E. Donnelly, P. Teeratchanan, C. L. Bull, A. Falenty, W. F. Kuhs, A. Hermann, and J. S. Loveday, "A Chiral GasHydrate Structure Common to the Carbon DioxideWater and HydrogenWater Systems," *The Journal of Physical Chemistry Letters*, vol. 8, pp. 4295–4299, Sept. 2017.

[62] B. Massani, C. Mitterdorfer, and T. Loerting, "Formation and Decomposition of CO_2-Filled Ice," *The Journal of Chemical Physics*, vol. 147, p. 134503, Oct. 2017.

[63] H. Hirai, T. Tanaka, T. Kawamura, Y. Yamamoto, and T. Yagi, "Structural Changes in Gas Hydrates and Existence of a Filled Ice Structure of Methane Hydrate Above 40GPa," *Journal of Physics and Chemistry of Solids*, vol. 65, pp. 1555–1559, Aug. 2004.

[64] J. S. Loveday, R. J. Nelmes, M. Guthrie, D. D. Klug, and J. S. Tse, "Transition from Cage Clathrate to Filled Ice: The Structure of Methane Hydrate III," *Physical Review Letters*, vol. 87, Oct. 2001.

[65] J. S. Tse, M. L. Klein, and I. R. McDonald, "Molecular Dynamics Studies of Ice Ic and the Structure I Clathrate Hydrate of Methane," *The Journal of Physical Chemistry*, vol. 87, no. 21, pp. 4198–4203, 1983.

[66] V. S. Efimchenko, M. A. Kuzovnikov, V. K. Fedotov, M. K. Sakharov, S. V. Simonov, and M. Tkacz, "New Phase in the WaterHydrogen System," *Journal of Alloys and Compounds*, vol. 509, pp. S860–S863, Sept. 2011.

[67] T. A. Strobel, M. Somayazulu, and R. J. Hemley, "Phase Behavior of H_2 + H_2O at High Pressures and Low Temperatures," *The Journal of Physical Chemistry C*, vol. 115, pp. 4898–4903, Mar. 2011.

[68] G. S. Smirnov and V. V. Stegailov, "Toward Determination of the New Hydrogen Hydrate Clathrate Structures," *The Journal of Physical Chemistry Letters*, vol. 4, pp. 3560–3564, Nov. 2013.

[69] T. A. Strobel, M. Somayazulu, S. V. Sinogeikin, P. Dera, and R. J. Hemley, "Hydrogen-Stuffed, Quartz-like Water Ice," *Journal of the American Chemical Society*, vol. 138, pp. 13786–13789, Oct. 2016.

[70] L. del Rosso, M. Celli, and L. Ulivi, "New Porous Water Ice Metastable at Atmospheric Pressure Obtained by Emptying a Hydrogen-Filled Ice," *Nature Communications*, vol. 7, p. 13394, Nov. 2016.

[71] V. Buch, R. Martoňák, and M. Parrinello, "A New Molecular-Dynamics Based Ap-

proach for Molecular Crystal Structure Search," *The Journal of Chemical Physics*, vol. 123, p. 051108, Aug. 2005.

[72] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Journal of Computational Physics*, vol. 117, pp. 1–19, Mar. 1995.

[73] M. B. Sevryuk, J. P. Toennies, and D. M. Ceperley, "Why are Para-Hydrogen Clusters Superfluid? A Quantum Theorem of Corresponding States Study," *The Journal of Chemical Physics*, vol. 133, p. 064505, Aug. 2010.

[74] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega, "A Potential Model for the Study of Ices and Amorphous Water: TIP4p/Ice," *The Journal of Chemical Physics*, vol. 122, p. 234511, June 2005.

[75] M. E. Tuckerman, J. Alejandre, R. López-Rendn, A. L. Jochim, and G. J. Martyna, "A Liouville-Operator Derived Measure-Preserving Integrator for Molecular Dynamics Simulations in the IsothermalIsobaric Ensemble," *Journal of Physics A: Mathematical and General*, vol. 39, no. 19, p. 5629, 2006.

[76] S. Nosé, "A Unified Formulation of the Constant Temperature Molecular Dynamics Methods," *The Journal of Chemical Physics*, vol. 81, pp. 511–519, July 1984.

[77] W. G. Hoover, "Canonical Dynamics: Equilibrium Phase-Space Distributions," *Physical review A*, vol. 31, no. 3, p. 1695, 1985.

[78] W. Shinoda, M. Shiga, and M. Mikami, "Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation under Constant Stress," *Physical Review B*, vol. 69, p. 134103, Apr. 2004.

[79] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant Pressure Molecular Dynamics Algorithms," *The Journal of Chemical Physics*, vol. 101, pp. 4177–4189, Sept. 1994.

[80] M. Parrinello and A. Rahman, "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, Dec. 1981.

[81] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*. Bristol [England] Adam Hilger, special student ed ed., 1988.

[82] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes," *Journal of Computational Physics*, vol. 23, pp. 327–341, Mar. 1977.

[83] H. A. Lorentz, "Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase," *Annalen der Physik*, vol. 248, pp. 127–136, Jan. 1881.

[84] D. Berthelot, "Sur la Dtermination Rigoureuse des Poids Moléculaires des Gaz en Partant de leurs Densités et de l'écart que celles-ci Présentent par Rapport á la Loi de Mariotte, Compt," *Rendus*, vol. 126, pp. 954–956, 1898.

[85] J. Hickman and Y. Mishin, "Temperature Fluctuations in Canonical Systems: Insights from Molecular Dynamics Simulations," *Physical Review B*, vol. 94, no. 18, pp. 1–10, 2016.

[86] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics, 2nd Edition | Thermodynamics | General & Introductory Mechanical Engineering | Subjects |*

Wiley. Wiley, New York, 1985.

[87] L. D. Landau and E. M. Lifshitz, *Statistical Physics - 3rd Edition, Course of Theoretical Physics Vol. 5*. utterworth-Heinemann, Oxford, 2000.

[88] Y. Mishin, "Thermodynamic Theory of Equilibrium Fluctuations," *Annals of Physics*, vol. 363, pp. 48–97, Dec. 2015.

[89] T. Ikeda-Fukazawa, S. Horikawa, T. Hondoh, and K. Kawamura, "Molecular Dynamics Studies of Molecular Diffusion in Ice Ih," *The Journal of Chemical Physics*, vol. 117, pp. 3886–3896, Aug. 2002.

[90] P. Geiger, C. Dellago, M. Macher, C. Franchini, G. Kresse, J. Bernard, J. N. Stern, and T. Loerting, "Proton Ordering of Cubic Ice Ic: Spectroscopy and Computer Simulations," *The Journal of Physical Chemistry. C, Nanomaterials and Interfaces*, vol. 118, pp. 10989–10997, May 2014.

[91] S. Mondal, S. Ghosh, and P. K. Chattaraj, "A Molecular Dynamics Study on sI Hydrogen Hydrate," *Journal of Molecular Modeling*, vol. 19, pp. 2785–2790, July 2013.

[92] A. Luzar and D. Chandler, "Structure and Hydrogen Bond Dynamics of Water-Dimethyl Sulfoxide Mixtures by Computer Simulations," *The Journal of Chemical Physics*, vol. 98, pp. 8160–8173, May 1993.

[93] J. Zhang, J.-L. Kuo, and T. Iitaka, "First Principles Molecular Dynamics Study of Filled Ice Hydrogen Hydrate," *The Journal of Chemical Physics*, vol. 137, p. 084505, Aug. 2012.

[94] N. J. English and J. S. Tse, "Mechanisms for Thermal Conduction in Methane Hydrate," *Physical Review Letters*, vol. 103, p. 015901, June 2009.

[95] S. N. Chakraborty and N. J. English, "Hydrogen-Bond Vibrational and Energetic Dynamical Properties in sI and sII Clathrate Hydrates and in Ice Ih: Molecular Dynamics Insights," *The Journal of Chemical Physics*, vol. 143, p. 154504, Oct. 2015.

[96] P. H. Berens and K. R. Wilson, "Molecular Dynamics and Spectra. I. Diatomic Rotation and Vibration," *The Journal of Chemical Physics*, vol. 74, pp. 4872–4882, May 1981.

[97] J. Li, "Inelastic Neutron Scattering Studies of Hydrogen Bonding in Ices," *The Journal of Chemical Physics*, vol. 105, pp. 6733–6755, Oct. 1996.

[98] L. Hakim, K. Koga, and H. Tanaka, "Novel Neon-Hydrate of Cubic Ice Structure," *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 1834–1838, May 2010.

[99] R. Kumar, D. D. Klug, C. I. Ratcliffe, C. A. Tulk, and J. A. Ripmeester, "Low-Pressure Synthesis and Characterization of Hydrogen-Filled Ice Ic," *Angewandte Chemie International Edition*, vol. 52, pp. 1531–1534, Jan. 2013.

[100] L. Hernández de la Peña, M. S. Gulam Razul, and P. G. Kusalik, "Quantum Effects in Ice Ih," *The Journal of Chemical Physics*, vol. 123, p. 144506, Oct. 2005.

[101] R. Ramírez and C. P. Herrero, "Quantum Path Integral Simulation of Isotope Effects in the Melting Temperature of Ice Ih," *The Journal of Chemical Physics*, vol. 133,

p. 144511, Oct. 2010.

[102] C. J. Burnham, Z. Futera, and N. J. English, "Quantum and Classical Inter-Cage Hopping of Hydrogen Molecules in Clathrate Hydrate: Temperature and Cage-Occupation Effects," *Physical Chemistry Chemical Physics*, vol. 19, pp. 717–728, Dec. 2016.

[103] D. Frenkel and B. Smit, eds., *Understanding Molecular Simulation (Second Edition)*. San Diego: Academic Press, second edition ed., 2002.

[104] P. P. Ewald, "Die Berechnung Optischer und Elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, pp. 253–287, Jan. 1921.

[105] T. Darden, D. York, and L. Pedersen, "Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems," *The Journal of Chemical Physics*, vol. 98, pp. 10089–10092, June 1993.

[106] L. Greengard and V. Rokhlin, "A Fast Algorithm for Particle Simulations," *Journal of Computational Physics*, vol. 135, pp. 280–292, Aug. 1997.

[107] J. V. L. Beckers, C. P. Lowe, and S. W. D. Leeuw, "An Iterative PPPM Method for Simulating Coulombic Systems on Distributed Memory Parallel Computers," *Molecular Simulation*, vol. 20, pp. 369–383, May 1998.

[108] `https://zenodo.org/badge/DOI/10.5281/zenodo.154672.svg`.

Acknowledgment

Alhamdulillah, finally, this work is done. It took me three years full of joy and sadness when having this study to accomplish. Only the Almighty Allah SWT who can make this possible to happen. Millions memorable memories have struck into my brain while being in Japan, doing many things that a PhD students obligatory proceed. Hopefully, what i have done here is the achievements of my parents long term dreams, in which i deeply realize this achievement would value not even a single water-drop of reciprocation compared to their whole universe sacrifice.

Million thanks and respect full for Prof. Shinichi MIURA who is very kind and very patient on supporting me performing academic research while studying. I am so lucky to have you as my supervisor. Million thanks also for Prof. Hiroshi IWASAKI who is always be there when i need help to overcome difficulties. Many thanks also to all of the Molsci members for always supporting me when doing research. Special thanks to Masato Ikezawa who help me a lot during my stay in Kanazawa and study trip to Osaka and Tokyo. My grate full to Prof. Mineo Saito, Prof. Manohide Sato, Prof. Fumiyuki Ishii, Prof. Tatsuki ODA and Prof. Kawaguchi for their academic support.

I also would like to show my gratitude to Directorate General of Higher Education, Ministry of National Education, Indonesia for the scholarship. The financial support gives me opportunity to be able performing study. I also grateful to Indonesian Student association for supporting me since the first time I came to Kanazawa. Special thank to Pak Dr. Abdul Wahid Yunus who provide me a shelter when I got no place to stay during the first week of our stay in Kanazawa; there was a miscommunication between me and apartment's office at that period of time. Many thanks to Mas Hasan Al Rasyid, Mas Isman Kurniawan and Mas Nurul Ikhsan for being my tutor in computer programming. Also, I want to deliver my grate full to Dinas Perikanan member who always becoming very best friends to share. Furthermore, million thanks to my DIKTI 2015's friends for the support, information and the true friendship ever shown in my life. I would also like to say thanks to Jurusan Fisika FMIPA Universitas Tanjungpura for being my academic home for the past 15 years. Many thanks also to all of my student in Indonesia who always supporting me, specially to the laboratory members. Since the first time i came here, I have noticed that my effort of teaching was not good enough. I am really sorry for the past year i failed to deliver the same academic atmosphere as Kanazawa University has gave to their student. But, as soon as i get home, i promise to deliver you the same atmosphere.

The last but not least, I would like to share my gratitude to my family in Indonesia. I

keep this best part in the last paragraph to describe the main reason why I came here. There are many aspects that I have to deduce while gaining more academic efforts. To further describe this, the most affected value would be the existence of myself among my lovely person to support. My small family composed of a beautiful wife and two lovely sons who had to abandoned since the first time i came here always put me in an encouraged path. I am really sorry for this to happen. No words worth to describe how i really love you all so much. Thanks to my family at Pontianak and Singkawang for taking care my Eka Hidayana, my Aditya Ammar Ghiffari and my Azka Emeraldy Atthaya during my stay in Japan. Thank you very much for the support. I dedicate this work to my Father and my mother.